THE COGNITIVE SCIENCE OF RELIGION:

A CASE FOR ATHEISM

_____

A Thesis

Presented

to the Faculty of

California State University Dominguez Hills

_____

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Humanities

_____

by

P. Wesley Edwards

Summer 2016

ACKNOWLEDGEMENTS

First and foremost, I would like to express my heartfelt gratitude to my thesis chair, Professor William Hagan, whose passion, expertise, and patient feedback did so much to shape the present work.

I would also like to express my deep appreciation to the committee members, Dr. Robert H. Cubillos and Dr. Jacqueline Shannon, for their time, patience, and support.

Finally, I wish to acknowledge the invaluable education and support I have received over the years from the staff and faculty at California State University, Dominguez Hills.

TABLE OF CONTENTS

PAGE

CHAPTER PAGE

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 thought and behavior, including its most counter-intuitive aspects. These unique features create fresh insight into certain contentious questions within the epistemology of religion. This paper focuses specifically on the CSR's epistemic significance for the rational justification of theism, and defends the view that CSR's findings significantly diminish this justification.

CHAPTER 1

INTRODUCTION

Belief in supernatural influences—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 least as long (Mithen 174-178).

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 other self-evident, directly perceived, or folk-intuitive beliefs, that is, as unreflective "givens." 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 is completely rational. As we will also see, others see in this same universality something explicable solely in naturalistic terms, and still others among them who see *dis*confirmatory evidence of God's (or gods') existence.

Recently, these naturalistic explanations have taken the form of evolutionary-based accounts of human cognition—accounts that purport to show innate biases, which strongly predispose humans to believe not only in supernatural agents, but supernatural agents whose characteristics are *constrained* in telltale ways. Interestingly, within the relatively new interdisciplinary field in which this debate is developing, a field known as

the cognitive science of religion (CSR), both the theistic and atheistic [1] camps *agree* not only on the underlying data, but also on much of the evolutionary framework used to explain it. Where they differ is in their philosophical interpretations of this field's 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 diminish 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 atheistic interpretation of CSR's findings (atheist-CSR objector arguments) and in the process survey the state of the field in the relevant areas of the philosophy of religion. As the chapter unfolds, we will discover Alvin Plantinga's profound influence on the shape and force of these objections. Chapter 4 will critically assess these objections, drawing on the recent work of a number of thinkers, and will conclude that a particularly under-developed area in this debate is in the application of recent work in argumentation theory relating to burden of proof and pragmatics, coupled with certain insights from the philosophy of science. Chapter 5 will apply these additional analyses to extend the critical assessment of the atheist-CSR objector arguments. This, It

---

[1] "Atheism" has some formal distinctions, one of which is between *positive* and *negative* atheism. The latter are commonly referred to as "agnostics," or sometimes, "non-theists." What all types of atheist share is an *absence of belief* in the Abrahamic God or any other supernatural god. See (Martin) for a throughout treatment of these varieties of atheism. For convenience, and given the focus of this paper, I will use the term "atheist" in its broadest, absence-of-belief sense.

will be argued, seriously undermines the rational justification 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. Accordingly, the following discussion is intended to introduce the key concepts of CSR along with some of the evidential reasons why this field is taken so seriously in the scientific community. However, rather than mounting a detailed defense of any version of CSR, this paper is concerned with assessing the epistemological consequences of CSR *if* it 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, Explaining Religion (Away?) 521). While this is arguably too broad, he follows it with an important clarification that bears directly on 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" and while the question of any god's existence is beyond the scope of CSR, it may nonetheless, "have epistemic implications for religious belief" (Jong, Explaining Religion (Away?) 522).

A more contextual, 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 (J. L. 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" (J. L. Barrett, CSR: What is It? 768). This makes clear that an understanding of CSR requires an understanding of one of its progenitors: cognitive science.

Significantly, cognitive science emerged as a reaction to *behaviorism*, the view that the mind is an indivisible, general-purpose processer that contains no "ready-made reaction[s]" or "innate ideas" [cited in] (Pinker 20). Heredity for the behaviorist simply has nothing to do with behavior, which is entirely formed by stimulus and response conditioning—by the *environment*. As such, the mind is thought of as a kind of "blank slate," the canvas on which is painted whatever social, political and religious teachings a child happens to be exposed to. Even her talent is environmentally determined. She *becomes* that painting. Culture is therefore *autonomous* from any inborn properties of the brain (Pinker 18-23). This perspective, still alive in the social sciences, can trace its roots directly to John Locke. Indeed, it was championed to fight racism and sexism, which were often justified by appeal to supposed innate limitations (Pinker 18).

Yet, this mind-as-blank-slate view has been on a collision course with a separate historical track. The walls that divide the human realm from the rest of the universe had been falling since Newton's time.[2] Newton showed that the laws on earth were the same

[2] I draw these examples, including the usage of the "walls" metaphor, from (Pinker)

as those in the heavens. Lyell bridged the dynamic, creative past with the supposedly static present. And several thinkers had breached different parts of the wall separating living and non-living matter, culminating in the biggest breach of all: Darwin's bottom-up, algorithmic explanation of life's complexity and *apparent* design. The remaining wall, arguably 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 here that both cognitive science and evolutionary psychology come on the scene, forcing open two breaches in this final wall.

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" (J. L. Barrett, CSR: What is It? 769). The ideas coming from 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 "software modules," also described as "inference systems" (Boyer 17), "cognitive domains [,] intelligences" (Mithen 37), or "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 variability in human cultural expression. Yet our shared cognitive architecture limits this variability in tell-tale ways. In language, for

example, our “grammatical programs” limit any language’s building blocks to verbs, subjects, objects, prepositions, etc. So, while there are *many* languages, they must all be built from the same cognitively-based elements (Pinker 37).

Crucially, these cognitive modules are considered “content rich,” which is to say that they are “pre-programmed” not only with certain problem-solving algorithms, 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-programmed, multiple-specializations view. It is a view that begins to shed light on why, for example, children can learn languages so quickly and unconsciously with very limited input (Mithen 44). But where did 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 showed 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 considerations applied to the development of the human *mind* would lead us to expect it to be more like a “Swiss army knife” than the undifferentiated, passive “sponge” of the blank-slate model (Mithen 43). Moreover, this “multi-mind” would not only be pre-programmed with rules and information, but would also be *proactive*, pushed by various goals and automatic cognitive responses (e.g., eagerly pattern seeking).

Why? Because these characteristics are far more adaptive than the passive, general-purpose behaviorist alternative. Consider, for example, the split-second, life-and-death decisions that creatures have to face on very limited information. A general-purpose, *un-biased* "computer" would have to analyze far too many variables and possibilities in order to assess the best course of action *in time*, while specialized pre-loaded knowledge and associated pre-programmed rules (all operating, for the most part, *unconsciously*) would produce (often enough) a survivable response much more quickly (Mithen).

The 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. 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 creativity, behavioral flexibility, and unpredictability would be extremely difficult to explain (Mithen 58-59). Therefore, if the modularity view is correct, then there is clearly *some* interaction between these modules.

One model of that cross-module interaction is that it is mediated by *another* module sitting on top of the rest—a meta-representational module (Mithen 59). Regardless of the specific model, the key point is that knowledge is understood to *flow* across these specialized modules allowing different cognitive tools to work on information outside their usual domains. One effect of this cross-domain data flow is 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 obsession with analogical reasoning and metaphor, but also, *art* and *music*.

Steven Mithen, for example, proposes that art is the result of cognitive fluidity between certain once-strictly separated modules. Art results from the fluidity that appeared between natural history module (interpreting indicators or "symbols" like hoof prints), social module (including communication), and technical module (producing artifacts from mental imagery). Once the walls between these modules became more porous, art arose as "artifacts/images with symbolic meanings as a means of communication" (Mithen 163). But if art could arise from the evolution of cognitive fluidity in humans, what of *religion*? Indeed, as an archeologist, Mithen notes that the archeological evidence suggests that the emergence of belief in supernatural beings and an after-life was concurrent with the emergence of complex art (around the Upper Paleolithic) (Mithen 174-178).

Using cognitive science and evolution to explain religion would not be the first attempt at a natural explanation of this uniquely human phenomenon. Natural explanations for religion go back at least as far as Xenophanes (570 – 480 BC) (Inwagen 128), who 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, Hume would open his book 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*" [italics added] (D.

Hume 14). Much closer to our own era, more specific "natural" theories of religion would emerge, such as those by Freud, Marx, and Feuerbach, all of which have been criticized, justly, as either too vague or too untestable (Inwagen). Yet, despite the fact that Darwin's initial insights were focused on the non-mental aspects of biology, the historical track beginning with his ground-breaking work and leading to modern evolutionary psychology began to suggest 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* (Guthrie) among others (Trigg and Barrett), followed somewhat later by works that emphasized the role of evolution, such as Boyer's *Religion Explained* (Boyer). It was during this period (2000) that Justin Barrett would introduce the term "The cognitive science of religion" (J. L. Barrett, CSR: What is It? 780), and characterize 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).

From this rich historical context modern CSR has inherited some of its key *cognitive* assumptions: the rejection of the "blank slate" in favor of one that sees the mind as a collection of collaborative, content-rich, innate cognitive functions that are independent of culture—at least in terms of their *basic* settings—and which are behind our well documented universal biases and predilections, including, for example, certain moral and social-exchange intuitions (more on these later). These specialized cognitive

modules are considered part of a *quick response* system, which operates largely outside of our conscious awareness or direction. Their rapid outputs, produced in response to specific stimuli, are generally interpreted by our conscious minds as *intuitions*, while our much slower reflective system takes these rapidly produced, subconscious deliverances as inputs for higher-level synthesis and evaluative reflection (Trigg and Barrett 5).[3] To understand the current CSR "Standard Model" we need to understand just what is it about this cognitive architecture that explains the ubiquity of belief in supernatural agency—that is, what puts "R" (religion) in CSR.

From an evolutionary perspective, religion can be seen as (1) a direct beneficial adaptation, (2) as a result of meme[4] selection, or (3) as a by-product—a *side effect* of the *normal* operation of cognitive faculties evolved for other purposes (Wood). It is worth noting that these evolutionary interpretations are not necessarily mutually exclusive. For example, while our receptivity to certain ideas and our predisposition to form ideas of a certain of type may be seen as side-effects of our cognitive architecture, once an idea *is* "captured" by our minds, regardless of its origin, its subsequent effects may be adaptive or maladaptive for the host (or for the meme). As we will see, however, it is the by-product interpretation that does most of the heavy lifting in CSR as it appears to best explain the data. It is, as a result, part of the Standard Model (Jong, Explaining Religion

---

[3] Evidence suggests such higher-level, conscious deliberation may be less like a balanced reasoner, and more like a "spin doctor", weaving stories to integrate these deliverances (Pinker 43).

[4] "Meme" is play on the word "gene" and is the idea (introduced by Richard Dawkins in *The Selfish Gene*) that pieces of information can use human brains as a Darwinian environment. Like genes, these memes evolve, not for the specific advantage of their hosts, but for their own survival (i.e., memes like genes are "selfish") (Dawkins, The Selfish Gene).

(Away?) 522). With this cognitive science and evolutionary psychology conceptual background in hand, let us see a typical way these have been deployed in CSR to *explain* religion.

A CSR Explanation for Religion

Humans, much more so than any other species, are freed from having to consider only those stimuli directly impinging on the senses from their immediate environment. We are able to *decouple* our cognition from the here and now, and entertain "what-if" scenarios. This is central to our ability to plan and to process information provided by others. In other words, our ability to decouple allows us to fire our "inference engines" (another way describing content-rich cognitive modules) *without* the actual presence of physical stimuli. This has many consequences, some unintended. As Boyer points out, "Supernatural concepts are just one consequence of the human capacity for decoupling representations" (Boyer 131).

Boyer goes on to argue that the combination of specialized inference systems and decoupled cognition may explain why so much of human behavior has no clear adaptive value. Persuasively, he uses the *universal* phenomenon of music as an example. Our uniquely evolved auditory cortex has areas that specialize both in pure tones and complex signal analysis (for language processing). Music under this view 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 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 music is a kind of *side effect* of capacities evolved for *other* purposes—capacities that are, as a result, susceptible to certain types of *artificial* or inadvertent stimulation.

Other examples include the near-universal *features* of human art, gadgets, and body decorations, which, for example, are almost always characterized by bilateral symmetry. As it turns out, bilateral symmetry signals good genes in a mate, as well as the effect of seeing a face that is looking directly at you—something highly relevant to our predator/prey modules (Boyer 133). Other relevant analogies should spring to mind at this point, such as opioids that *happen* to fit receptors designed for other internally produced molecules, and artificial enhancements to sexual characteristics (e.g., 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 parent's beak. Experiments show that if a stick with not one, but *three* red stripes, are presented to the gull chicks, they will peck themselves into a bleeding frenzy. This artificial situation stimulates their mental wiring much more intensively than would a natural beak. As Ramachandran puts it, "If herring gulls had an art gallery, they would hang a long stick with three red strips 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).

Whether such artificial super-stimuli are deliberately concocted, or stumbled upon in nature (such as with opioids), is quite beside the point. The 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 a flame or lightbulb can lead a moth to spiral to its death (Dawkins, The God Delusion 201). For much of the moth's evolutionary history, bright light like the sun and moon, would have been the most frequently encountered light sources. Since these are at optical infinity, a simple rule like "keep it to the left," could have had selective advantage.

Supernatural concepts also strongly stimulate our inference systems. As a result, they grab our attention, and so are more likely to be mentally rehearsed, retained, and transmitted; however, unlike the examples cited above, some supernatural concepts—in particular, "religious" ones—activate those inference systems with the most intense emotions, particularly those related to *social interaction* and *morality* (Boyer 135). More than simply being fascinating, they *matter*. Crucially, our 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* existence, or the workings of the universe, in some *abstract* sense. What we typically encounter are supernatural concepts that are immediately *practical*, and are comprised of entities with which one can *deal*. These supernatural "agents" are invariably represented as having human-like *minds*, complete with goals, intentions, and memories—entities with which one can *interact*. In other words, they both *matter*, and 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. 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 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 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* them. As 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.

The fact that supernatural agents are *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" (D. 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 sophisticated: they adopt Dennett's *intentional stance* (Dennett 109-110), a term he uses in preference to "Theory of Mind" (ToM). This is an inference system(s) that *eagerly* (pro-actively) detects intentional *agents*. "Agent" 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 makes them socially relevant, particularly in "morally-laden" (Clark and Barrett, Reformed) areas of human social interaction, which relate to such inborn emotions as obligation, trust, and loyalty, which 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 HADD, or *hypersensitive agent detection device* (J. L. Barrett). 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 be far more damaging than would 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 engage those mental modules designed (pre-programmed/pre-loaded) to respond accordingly, for example, deception, submission, creating obligation by giving something of value, etc. If one wrongly "detects" an intentional agent where none exists, then some energy is expended, 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 is *much* higher: death by predation (from inside or outside the species), or being outmaneuvered in the social hierarchy (through deception, cunning, etc.). This HADD is precisely what leads us

to overdo things—to project intention where none exists, such as on the weather and other random events.[5]

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--animal-human hybrids (totemism) (Boyer), conscious mountains, trees with memory, etc.—our nearest relatives, despite evidence for some decoupled cognition, show no signs of such cross-category fantasizing or mythical/religious thinking (Mithen). It is important to note that while we, as a species, have indeed 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 (or goes "viral," to use a modern metaphor). 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 many implicit (subconscious) inferences that are automatically applied to any newly hypothesized *concrete* category instance (Boyer 60-61). Those concrete concepts with attention-grabbing staying power are those that violate some aspects of an ontological category, but not so many that the category no

---

[5] HADD, in combination with our other inference engines, might give some insight into both the diversity and baroque complexity 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. No matter what the pigeon was doing, 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 cause-effect-seeking inference engines were linking their at-the-moment behaviors to *random* events (Dennett 118).

longer applies. That is, they are *minimally counterintuitive*. By preserving the category, all the implicit expectations produced by it remain. This is what helps separate the fascinating from the just plain weird: *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 inferences; and (2) add some special category-violating features, typically with important social implications. This kind of combination stimulates, indeed captures, our fascinated attention. It is, for humans, like the three red stripes on a stick. Boyer cites one example involving 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, this "juicy gossip" 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 cognitive powers like "knows all") (Boyer 63-64).

A final key CSR concept is something Justin Barrett describes as Theological Correctness or TC (J. L. Barrett, CSR: What is It?). 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 there are more than a few courses of action open to Him (assuming for the moment that the notion of "taking action" is even coherent in the context of *omnipotence*). 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 such time that they are rescued. The results of this experiment were 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 140-141). Boyer observes that if God is a Person category instance, then our spontaneous intuition (i.e., unconscious, rapid inference from a specialized module) is that it is far easier for a *person* to change someone's mind than it is for him or her to intervene in the natural processes 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 is activated only because people represent God as a person-like agent who *interacts* with them" (Boyer 142). Note how this relates to our earlier observation about *theology* not even being part of many world's religious practices.

The CSR Standard Model

With these concepts in hand, we can now describe the CSR Standard Model with regard to beliefs in supernatural agency: [6]

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

- They are the outputs of cognitive systems evolved for detecting agency. This is where HADD comes in.

- In addition to being intentional agents with their own desires and beliefs, sometimes referred as Theory of Mind (ToM), supernatural agents are socially relevant. We can enter into exchange with them, and they are able to help us in our dealing with other members of our group.

- To the extent that there is an espoused theology—a sense of Theological Correctness—this appears to break down in practical applications where our category-based anthropomorphic inferences assert themselves.

One consequence of this pre-dispositional / by-product view is that we expect our cognitive systems to produce not *specific* supernatural beliefs or theologies, but rather a general *susceptibility* to supernatural beliefs that satisfy our aggressive HADD-based impulses while meeting the conditions described above.

On a conceptual level, CSR's empirical approach has the interesting consequence that it allows work to proceed without a fully fleshed-out definition of "religion." This is because it investigates ostensible religious phenomena on a case-by-base or "piecemeal"

---

[6] This is adapted from (Leech and Visala) and (Murray and Goldberg, Evolutionary Accounts)

basis (Trigg and Barrett 6).[7] As we will see later, this "bottom-up" approach resonates with a 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's and Barrett's definitions 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 note that even if CSR were to accurately explain the universality 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 (or any other natural) terms "explains it away" is a philosophical question whose answer must be argued and not assumed—and its answer is by no means obvious. For example, Justin Barrett is not only a Christian, but a major,

---

[7] Barrett and Trigg claim that, philosophically, this limits the impact of epistemic problems with religious belief since it may be limited to that *particular* religious belief rather than all of them (Trigg and Barrett). We will later identify reasons to believe that the negative implications for the rational justification of theism are indeed more general than they claim.

founding scientific contributor to this new field and someone who has argued forcefully for a theistically compatible 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. As we explore the recent state of this discussion, we will also review the state of the field in relevant areas of the epistemology of religion.

CHAPTER 3

OBJECTIONS TO AN ATHEISTIC INTEREPRETATION 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 atheistic arguments (by atheists) based on CSR have been rare; instead, the field's purported atheistic implications have more 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" (J. L. Barrett, Is the Spell 60). Similarly, Murray and 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, Evolutionary Accounts 194).[8] 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 to atheist-CSR, have often had to first formally develop the

[8] Boyer, for example, famously likened supernatural belief to poetry, in that they both give to "airy nothing a local habitation and a name" (Boyer 4). See also (Atran).

atheists' implicit arguments.[9] Doubtless, some of this is due to the fact that CSR was pioneered by scientists—anthropologists, archeologists, and cognitive scientists—rather than philosophers.

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—among his many other noteworthy contributions—have deeply influenced the philosophical discussion surrounding CSR. As a result, reviewing his insights will be an important first step toward understanding the history and context of the current philosophical debate regarding whether, and to what degree, CSR's findings bear on the rational justification of theism.

Alvin Plantinga's Foundational Work

Critique of Classical Foundationalism

Hume's famous evidential dictum, "A wise man . . . proportions his belief to the evidence" (D. Hume), has been cited and paraphrased in time and again in critical thinking discussions. Many atheists are quick to point to this principle when explaining their insistence that belief in supernatural beings should *at a minimum* be withheld (the "negative atheist" position) due to a lack of sufficient evidence. In a similar vein, but

---

[9] See for example (Murray, Scientific).

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 inference from other inferentially justified beliefs? If so, then it would appear to be vulnerable to the charge of *reductio ad infinitum*—an infinite regress. As it so happens, however, evidentialism comes in various forms, including those that are not vulnerable to this charge, such as coherentism (Berker). For others, the terminus of the justificatory chain is some form of *foundationalism*, which holds that certain beliefs are *basic*—that is, can be rationally held without *independent* justification.

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 (henceforth 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,

$$(1)\ P_{bs} \leftrightarrow\ P_s \wedge P_i \wedge P_e$$

Plantinga understands "self-evident" here in the sense that as soon as one grasps the proposition, one *immediately* (as in *un*mediated by other propositions) sees that it is true.[10]

He identifies two senses of this immediate grasping of basic beliefs: epistemic, and phenomenological. The epistemic sense captures the above discussion about the *immediate* apprehension of a proposition's truth; the second seems to be his effort to capture the subjective, qualitative experience associated with the irresistible need to *accept* a proposition's truth—a kind of "aura" uniquely associated with the recognition of the truth of properly basic beliefs, and absent from non-basic beliefs (Plantinga and Wolterstorff, Faith and Rationality 57).

In clarifying the idea of being "evident to the senses," he distinguishes what he understands to be Aquinas' less tentative sense from Descartes' more qualified sense. 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—evident to the senses—Descartes would arguably insist that only the

---

[10] It is worth noting that what is immediately apprehended by one person may not be to another. For example, to the extent that one does not have to do derivational computation to convince oneself of the truth of the statement "1 + 1 = 2," the statement is properly basic for that person. For many of us, the truths of more complex mathematical statements are not properly basic precisely because awareness of their truths is *not* immediate, though they may become so with training.

certain, subjective *seeming* to see a tree would qualify as truly basic. It is in this latter sense that we mean a proposition is *incorrigible* (Plantinga and Wolterstorff, Faith and Rationality 58) [11].

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, (2) and (3):

$$(2)\ P_s \wedge P_i \wedge P_e \longrightarrow P_{bs}$$

$$(3)\ P_{bs} \longrightarrow P_s \wedge P_i \wedge P_e$$

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, and it is with this that Plantinga takes issue.

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. 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, nor evident to the senses. Neither is this belief derivable from other basics beliefs. Nonetheless, it *is* basic to the person with the memory, and this

[11] For an excellent broader treatment of philosophical incorrigibility, see (Raff)

hardly seems irrational, even if it is, under certain circumstances, *unreliable* (Plantinga and Wolterstorff, Faith and Rationality 76).

Plantinga's most powerful charge, however, is that the classical foundationalists' central principle is *self-refuting*, at least until such time that independent support for it can be found. Such support, however, would have to meet the very conditions that (1) *itself* lays out. Without such support, Plantinga asks, on what basis are we to accept (1), since it is not basic by *its own criterion*, nor is it derivable from other properly basic propositions. In other words, to accept (1) is to violate it (Plantinga and Wolterstorff, Faith and Rationality 60).

Plantinga is rejecting just one particular classical foundationalist condition, namely, that for a person S, only those beliefs that are either incorrigible, self-evident, or evident to the senses, can be accepted as properly basic (that is, if S's noetic structure is to be considered *rational*). What he terms the Evidentialist Objection to belief in God is the charge that belief in God fails to meet this condition. However, through the above critique, he believes that he has defeated this objection by defeating the rationality of CF on its own terms. By arguing that this common justification for rationality is, ironically, irrational, he opens the door to what he hopes is a much better alternative approach—one that puts theism on a defensible *rational* footing. He does this by arguing that belief in God is, in fact, properly basic.

Defense of Theism as Justified Properly Basic Belief

Plantinga makes an important clarification regarding his notion of a "basic" belief. Beliefs about our own internal mental states, such as "I see a dog," while basic, are not (necessarily) *groundless* or *unjustified*. The justification for such beliefs 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 asserts that there are just such justifying preconditions for rational, *prima facie* belief in God. Here, Plantinga 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, WCB 170). Calvin expanded this into the notion that God created us with a predisposition to believe that various actions and objects are the result of God's handiwork—a *sensus divinitatus* (Plantinga, WCB 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 function* of the *sensus divinitatus*. He asserts too that we seem similarly disposed to *relational* beliefs involving God, which produce feelings of gratitude and guilt (Plantinga, Is Belief in God Properly Basic? 46).

This view that belief in God is properly basic is also referred to as *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, WCB). His view that belief can be justified on the basis of a cognitive process's reliability (in the sense of its "truth conduciveness") coupled with facts *external* to the believer's mind *that the believer may not have access to* is referred to as *reliabilism*, which is an instance of *externalism*. In contrast, *internalism* generally means that a believer does (or can) have access to the basis for her belief's justification. This distinction bears on Plantinga's particular brand of reliabilism, known as *proper functionalism*, since it is mainly a theory of warrant rather than justification. We can understand this distinction by recognizing *justification* as that subtype of warrant that derives from reasons the believer can access. In other words, justification is internalist warrant (Altschul). Given that reliabilism arose in the 1970s and 1980s as a reaction to evidentialism (in mainstream epistemology), it is generally considered

incompatible with evidentialism.[12] It is evidentialism that is often combined with classical foundationalism in order to avoid the infinite justification regress objection we saw earlier.

Plantinga's attempt to undermine classical foundationalism and establish a rational basis for justified belief in God has been challenged as opening a Pandora's Box of "justified" beliefs, which he dubs The Great Pumpkin Objection (Plantinga and Wolterstorff, Faith and Rationality 74). Why, in other words, can one not declare belief in the Great Pumpkin to be a properly basic belief? Plantinga's response is interesting. He maintains that establishing *hard* criteria for proper basicality is arguably impossible since it can be approached only bottom-up, rather than top-down. Such a bottom-up approach would necessarily be an inductive one, for which he suggests the following procedure. First, assemble a relevant set of basic-belief examples, that is, clear, not-so-clear, and even 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, Faith and Rationality 76). Note that such a scientific-like approach means that we would have to be prepared to revise our belief-condition example sets based on those test results. He concludes, therefore, that the most

[12] Though some have argued otherwise [see for example] (Smith)

we can hope to achieve is establishing criteria for *prima facie* (and therefore defeasible) justified basic belief (Plantinga and Wolterstorff, Faith and Rationality 77).

What are the implications of this 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 be the criteria 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, Faith and Rationality 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, Faith and Rationality 78). But this is also why the Christian is epistemically within her rights to reject The Great Pumpkin as properly basic. Note that he 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 early 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 believes that God did, in fact, *design* P to be truth-tracking with regard to religious beliefs. *If* Plantinga, and like-minded theists are right about this, *then* (of course) B would be warranted.

On logical grounds, this appears to be an uninformatively trivial truth, comparable to "If I am hungry and the sky is blue, then the sky is blue." As Dawes and Jong observe, the form of his argument is simply

$$A \wedge B \longrightarrow B$$

where Christian beliefs can be represented as a long conjunction. Here, A is the set of Christian beliefs exclusive of Christian beliefs about warrant, and B is the remaining set of Christian beliefs *about warrant* (Dawes and Jong 2). But they note too that this triviality, when *combined* with the idea that Christian beliefs are *basic* with respect to warrant, produces implications that are less obviously trivial.

As discussed earlier, if a belief is basic, it is not inferred from other beliefs, but is "immediate" in the sense of being the spontaneous effect of certain experiences in the context of relevant justifying 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 justified *basic* belief, it is neither based upon evidence, nor warranted *because* it is based upon evidence. Consequently, it is not subject to disconfirmatory evidence in the same way that non-basic beliefs are, such as belief in a scientific theory. To illustrate the force of this, Plantinga uses the example of an innocent person convicted by genuinely overwhelming (public / intersubjectively verifiable) evidence that he slashed someone's tires (Plantinga, WCRL 359). Despite this evidence,

however, the accused clearly *remembers* spending the afternoon elsewhere busily *not* slashing tires. That memory-based belief is basic for this person irrespective of the evidence, and its basicality is what *rationally justifies* his belief in his own innocence despite the strong evidence.

While the admittedly overwhelming evidence is understandably persuasive to the jury, it is also understandably *un*persuasive to the accused. In other words, the accused's edifice of mutually supporting beliefs—his noetic structure—includes one important thing that the jurors' lack: one particular properly basic belief that acts almost like a *singularity*, which overwhelms in a negative direction the guilt-positive effects of all other evidential considerations. Plantinga's tire-slashing analogy could be used to explain how certain pieces of evidence against Christian theism can, from an atheist's perspective, seem to be mystifyingly, even maddeningly, unpersuasive to thoughtful believers[13].

In building this case, Plantinga explains his particular notion of rationality. First, it is contingent upon the notion of proper function, which is to say that one's belief cannot be rational if there is *dysfunction* in any of the relevant faculties. Second, there is 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

---

[13] Dawes and Jong point to the example of the Argument from Evil (Dawes and Jong 3). Coupled with various Old Testament passages, this can indeed seem to be a more-than-decisive case against one of the Christian God's core attributes: His *goodness*.

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. Note that *internal* rationality, when functioning properly, produces beliefs that cohere logically with that person's *other* beliefs.

To illuminate these ideas, he points out that someone who truly believes that she is made of glass may well be *internally* rational (Plantinga, WCB 112). Although her external rationality is not functioning properly, as evidenced by the inappropriate *seeming*, her internal belief-forming faculty may nonetheless be responding normally *in response to* that experiential input, such as by producing a belief that coheres with her other beliefs. Significantly, this internal/external distinction allows us to *uncouple* belief formation from the present moment. This is why, for example, we can have doxastic experiences associated with memories and *a priori* beliefs (Plantinga, WCB 110-111).

Since theistic (especially Christian) belief is warranted provided that it is true (since if it is true, then the mechanism is functioning properly when it delivers belief under appropriate conditions), then the burden is on the atheist to show that it is untrue., In meeting this burden, one can object to Christian belief in one or both of the following ways: asserting that its content is false, and/or asserting that it is unjustified *regardless* of

its content's truth, perhaps by arguing against proper function. This is his *de facto* versus *de jure*[14] distinction.

So far in our discussion, Plantinga has worked to first create a space for rational, justified theistic belief by critiquing classical foundationalism. Then he 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 argument is part of Plantinga's development of Reformed epistemology. But he is not content to retreat to the safety of these new theistic fortifications. He then goes on the attack, arguing that without theism, an atheistic (i.e., naturalistic) interpretation of one of science's most successful theories—biological evolution—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 this paper. To fully appreciate this argument, it is important to understand its scope. Out the outset, he 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

[14] We will, along with Plantinga, and strictly in the interest of space (not necessarily agreement), focus on the rational-justification sense of *de jure* rather than on the deontological sense. He dismisses such deontological-based objections as "much too easy to answer" (Plantinga, WCB 108).

his example, *all* of one's beliefs *could* be false, but one could not sensibly believe this to be the case (since, of course, *that* would be one of the beliefs) (Plantinga, WCRL 310).

In building his EAAN, Plantinga applies this basic logical point to the presumption that our belief-producing faculties can be sensibly believed to be reliable *given that* they were *solely* the product of evolution in the context of *naturalism*, which, for a (naturalist) evolutionist, entails that evolution has no foresight, and has been completely unguided. (Plantinga, WCRL 311). Plantinga assumes that naturalism entails materialism, which is to say that it entails, among other things, the belief that humans are nothing more than their physical parts (Plantinga, WCRL 318). Or, in Gilbert Ryle's famous phrase, that there is no "ghost in the machine." More generally, naturalism assumes that there are no non-natural (e.g., supernatural) entities. [15] 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, WCRL 316).

---

[15] There is an important distinction between metaphysical and methodological naturalism, where the latter does not make assertions about whether supernatural entities do or do not exist. Rather, as a *methodological principle*, it precludes incorporating such notions into scientific explanations. There is, for example, no tolerance for incorporating into a scientific explanation something like, "at this step a supernatural event occurs." Plantinga's argument does not seem to depend on this distinction, however, as evolutionary theory entails only that the process is in no way *purposeful*, regardless of whether such supernatural entities exist.

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 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, WCRL 317).

Why should we believe Premise 1? Plantinga claims that evolution can indeed be expected to produce adaptive *behavior* geared toward enhancing survival; however, behavior and *belief content* 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 being food. But, he claims, *beliefs* need not enter into this at all. (Indeed, some creatures, like the jellyfish, almost certainly have nothing like “beliefs,” at least in any conscious 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 (captured via neural indicators, beginning with the sense organs) (Plantinga, WCRL 326-328). This disconnect between behavior and belief-content means, according to Plantinga, that belief content is not subject to evolutionary pressures one way or the other with the result that beliefs are as likely to be objectively true as false (Plantinga, WCRL 330-331). Plantinga seems to be

arguing that the naturalist has no reason to see belief content as 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, WCRL 331).

This disconnect between belief content and behavior leads us directly to the issue of cognitive *reliability*. Regardless of how we 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 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, WCRL 332-334).

Premise 2: "Anyone who accepts (believes) N&E and sees that P(R/N&E) is low has a defeater for R" (Plantinga, WCRL 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, WCRL 359). In other words, if you come to believe a defeater, say B_d, for belief B, then rationality constrains you to either believe its denial (~B) or to be agnostic toward it (Plantinga, WCRL 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, WCRL 345). This requires little elaboration, as 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 and can't rationally be accepted." Therefore, "N&E can't be rationally accepted" (Plantinga, WCRL 345).

The conclusion means that the atheistic, naturalist perspective, along with evolution, leads to what Plantinga calls a "crushing skepticism" (Plantinga, WCRL 345). After all, if *all* of our beliefs, including our beliefs in N, E, and even *basic logical principles*, are deliverances of the very cognitive faculties now deemed unreliable, then argument of *any* kind—including arguments related to the reliability/unreliability of our belief-forming faculties—cannot even get off the ground. Any such argument would *presuppose* that there is sufficient reliability in those same faculties to be able to form and understand a *coherent argument*.

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 convincingly 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 even R ⊨ (entails) D, then defeat of *any* belief would become impossible, since P(R/N&E&R) would always be 1. (Plantinga, WCRL 349).

Summary

Plantinga has carefully attempted to create a rational, epistemological foundation for theistic belief. His impressively methodological 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 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. The influence of these ideas on the shape of the philosophical debate surrounding the justification of theism given the emergence of CSR is difficult to overstate. Justin Barrett, one of CSR's founding contributors, and others, build directly upon this work.

Barrett's and Church's EAAN-Based CSR-Specific Critique

Justin Barrett and Ian Church leverage Plantinga's EAAN in an effort to undermine any atheism-supporting interpretation CSR's findings, defend theism, and bolster the EAAN itself by providing it with empirical support. To appreciate their analogical approach, 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,[16] in relative terms, the numerical disparity between atheists and theists is enormous.

With this in mind, they build their argument on an analogy. Given atheists' tiny-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 that of her many companions. Of course, 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' interpretation 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 that the non-reflective beliefs that these tools steer us toward are theologically *incomplete* (Barrett and Church, Should CSR 3), which would be consistent with the plethora of religious and supernatural concepts we now observe in the world.

As we have seen, the *normal* operation of our evolved cognitive tool suite evidently involves a very sensitive ("hypersensitive") agency detector, HADD, which is tuned to reflect the low-cost of false positives and the high (even fatal) cost of false

---

[16] 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 persons or a convinced atheist" (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, even a small fraction of the WIN-Gallop result would be far from insignificant.

negatives (e.g., it is generally better [cheaper, if you will] to mistake a branch for snake than a snake for a branch). One side-effect of this cost-asymmetry is the human tendency to over-attribute agency, which leads to the attribution of supernatural agency when natural agency is not readily apparent. Are these false-positive outputs of HADD analogous to our partygoer's friends' seeing the other guests through alcohol-distorted lenses?

Barrett's and Church's response is insightful. While we have independent reasons for thinking alcohol negatively affects the reliability of our senses, what similarly *independent* reasons do we have for supposing that our belief in the supernatural is similarly unreliable? They argue that once we remove the atheist's *prior metaphysical commitments,* we would see that there are none. After all, how can we assess the reliability of our god-belief-generating faculties unless we have *independent* knowledge of the presence or absence of such entities when these BFFs are in operation? (J. L. Barrett, Is the Spell 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, Should CSR 5-6). 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 are *always* using the same BFFs—we never "sober up," as it were (Barrett and Church, Should CSR 16).

The analogy fails in another way too: Our people-related BFFs produce fairly detailed perceptions of the other guests in all but the most extreme cases of intoxication.

However, our religious BFFs produce only the vaguest perceptions of the divine. 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. Why would this natural knowledge of God not be more specific? Plantinga's answer is that it "has been compromised, weakened, reduced, smothered, overlaid, or impeded by sin and its consequences" (Plantinga, WCB 184). But, Barrett and Church claim, this does not matter to their argument. Regardless of how vague such deliverances may be, for the theist, they at least point in the generally *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 just false, but false on such a global and historical scale that they 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* with regard to religious belief, the atheist is committed to the idea that this mechanism is *unreliable*. 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 atheist's relevant set of commitments, given CSR:

(1) There are no gods, souls, and afterlife.

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

(3) Hence, CSR-BFFs typically produce false beliefs.

(4) BFFs that typically produce false beliefs are unreliable.

(5) Hence, CSR-BFFs are unreliable.

(6) Beliefs formed by unreliable faculties lack warrant.

(7) 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.

(8) These beliefs (from 7 and 5) lack warrant (Barrett and Church, Should CSR 7).

If this argument holds, then the atheist is forced into skepticism toward *all* her beliefs, not just religious beliefs. This leads to 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 without being able to specify the differentiating conditions between the two cases, such a move would amount to special pleading (Barrett and Church, Should CSR 8). On this point, they cite 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, Scientific 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 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 in order 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, Should CSR 15).

Jong's and Visala's Bayesian Irrelevance Argument

As we have seen, Plantinga treats the notion of a reliable BFF as one that produces significantly more true results than false. Jong and Visala take a closer and more formal look at the notion of reliability, and at how it is used in arguments that purport to use evolutionary explanations (not just CSR) to undermine theism, so-called Evolutionary Debunking Arguments (EDAs) (Kahane).

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 *because* the belief is unjustified (Jong and Visala, EDAs 245). Second, they analyze the notion of a BFF's being reliable in terms of whether or not the process is "off-track," a concept built upon 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 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) [cited in] (Jong and Visala, EDAs 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.

In this context, we can recognize Boyer's as an example of an EDA that seeks to debunk theism solely by explaining theistic belief formation in entirely naturalistic terms. But is that enough? As we have seen, Barrett and Church have argued that this works only if one imports certain naturalist metaphysical assumptions. Jong and Visala, as we will see, go further and point out that, absent these assumptions, a natural explanation is not *inconsistent* with theism since a complete natural explanation would be explaining *one* sufficient cause but not necessarily the *only* sufficient cause. Moreover, other causes might exist not only as other sufficient *proximate* causes, but also at higher levels of explanation.

Jong has argued elsewhere (Jong, Explaining Religion (Away?) 528) that, like scientific explanations generally, religious explanations also occur at different levels, including at the *ultimate* level. 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 sits above and *subsumes* all other explanations, including *naturalistic* ones. In such a case, there is no necessary conflict between religious belief and a purely scientific explanation, since the latter operates at lower explanatory levels. In the case of CSR, a theist could 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, then, scientific evidence simply does not bear on the question of God's existence or on His role as ultimate designer.

Assuming for the moment that the burden of proof is on the skeptic (an assumption we will later challenge), it seems that we need more than one sufficient natural explanation for theistic belief in order to debunk its justification. Perhaps we need to show that the responsible belief-forming process is *more likely* to produce false beliefs than true ones, which is to say *given that* one believes p (through some BFF), then the probability that p is true is actually *lower* than the probability that *not*-p. This can be expressed[17] in conditional probability terms as

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

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

[17] What follows is an inequality-based variation of Jong's and Visala's use of Bayes Theorem, which, I hope, helps to make their point even clearer. The analyses are, of course, equivalent in their results.

Given that the prior probabilities are understood to include all background information, Bayes' Theorem[18] allows us to analyze the preceding two conditional probability terms as follows:

$$\mathrm{P}(\sim \mathrm{p} \mid \mathrm{S\ believes\ p}) = \frac{\mathrm{P\ (S\ believes\ p} \mid \sim \mathrm{p}) \ \times \ \mathrm{P\ (\sim p)}}{P(S\ believes\ p)}$$

$$\mathrm{P(p \mid S\ believes\ p)} = \frac{\mathrm{P\ (S\ believes\ p} \mid \mathrm{p}) \ \times \ \mathrm{P\ (p)}}{P(S\ believes\ p)}$$

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

$$\frac{\mathrm{P\ (S\ believes\ p} \mid \sim \mathrm{p})}{\mathrm{P\ (S\ believes\ p} \mid \mathrm{p})} \ \times \frac{\mathrm{P\ (\sim p)}}{P(p)} > 1$$

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*, then the left-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

$$\frac{\mathrm{P\ (\sim p)}}{\mathrm{P\ (p)}} > 1$$

This means that in the uncorrelated case, debunking p amounts *only* to showing that the *prior* probability that p is false is greater than the *prior* probability that p is true. But that

[18] For an overview of Bayes Theorem, see the Bayes Theorem Explained section, below.

means that considerations associated with the belief-forming process become *completely irrelevant.* The whole question simply falls entirely back onto the philosophical arguments we have historically used, with no contribution at all from CSR (Jong and Visala, EDAs 247).

This Bayesian analysis helps formally capture a central idea behind the Genetic Fallacy: *how* one comes to believe something—indeed, belief itself—does not bear on its objective truth. (As we will see shortly, however, this is not true in all situations.) But what about the *justification* for that belief regardless of its objective truth? This is, of course, a different matter. For example, if a theist appeals to a non-truth-tracking belief-forming process as the *sole justification* for her belief, then she would indeed be unjustified in holding that belief, even if the object of her belief were true (Jong and Visala, EDAs 255).

Despite the Genetic Fallacy worry, a similar Bayesian analysis can show how there actually *can* be a connection between a belief-forming process and the likelihood that the object of that belief is true. This is the *negatively* correlated case. In this situation, instead of the left-most ratio becoming "1," it would become something greater than 1. Indeed, it could become as large as appropriate depending on how "perniciously deceptive" (Jong and Visala, EDAs 248) we think the belief-forming process is. Consider the following situation: 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's and Visala's example of .9 and .0001 we would get

$$9000 \times \frac{\mathrm{P}\,(\sim \mathrm{p})}{P(p)} > 1$$

This would mean that the prior probability for p's truth could be much higher than the prior probability of its falsehood, and 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're true), but until then, we need not worry about EDAs that posit falsehood-tracking or perniciously deceptive cognitive processes" [italics in original] (Jong and Visala, EDAs 249).

We will examine in a later section some reasons to suspect that they have dismissed this scenario too quickly[19]. For the moment, however, let us accept their assessment of the non-applicability of the negative correlation case and return to the particular uncorrelated case where a believer's *sole* justification for belief is the BFF itself. Does such a situation describe actual theists? Interestingly, it does: it seems to correspond to epistemological *reliabilists*, including Reformed epistemologists. This group *justifies* their belief on these process' presumed truth-tracking abilities. Recall that Reformed epistemologists explicitly reject evidentialism. To be sure, evidentialists may, along with reliabilists, agree that belief was *originally* caused by non-evidential sources,

[19] 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. See, for example (Dawes and Jong 8-9).

such as religious experience, but it is the evidentialists who *justify* their belief on the 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, EDAs 256). Later we will argue that this vulnerability can be expanded.

Thurow's Propositional-Doxastic Distinction

Jong's and Visala's relevancy critique is powerful, yet limited in scope. Josh Thurow takes a different approach, which can be interpreted as expanding that scope. In developing his case, Thurow first works to clean up what he sees as important vulnerabilities in the current critiques of the atheist-CSR position. His approach is to iteratively develop and critique his understanding of that position. He begins by proposing an *initial* 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 point 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, 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's and Visala's point 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: "If there were no gods and we were to form [and wholly sustain] beliefs about gods using process P, we would still believe, via P, that some kind of god exists" (Thurow 86-87). His point is that if P is a *basic-belief* forming process and not an inductive process, it would seem that P is indeed unreliable with regard to belief in god. One important aspect of this particular reliability test is that it does not require that we know prior to using our faculties that they are reliable, only that we have evidence that

the faculty under consideration would produce a false *basic* belief *even if* the object of that belief were false (Thurow 89).

This would, initially, seem to address Plantinga's, Barrett's, and Church's worry about naturalist question-begging, both in terms of presuming atheism, and in presuming the unreliability of our cognitive faculties in 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. Such a device is "reliable," according to Thurow, because the device's *designer* based it on principles she knew to be true[20]. Similarly, if God did in fact design our CSR-based BFFs, then they would be reliable in precisely this sense, despite any current inaccurate outputs. His point is that just because a process produces an incorrect result, it does not necessarily mean it is "unreliable," which undermines P2.

Yet, a further thought experiment suggests 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 belief 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

---

[20] Although Thurow does not explicitly say so, this device must also be reliable in terms of its *mapping* between inputs and outputs—that is, one must be able to count on not getting a randomly *different* answer with the *same* inputs. In other words, the device must be *internally* reliable.

science of Martians" discovery should lead us to suspend belief in life on Mars pending independent evidence of its existence. Notice that our belief in life on Mars may have been *prima facie* justified, but the new discovery created a *defeater* for that initial justification. We should, therefore, suspend belief in life on Mars pending evidence, but if we find any, then it may lend strong support to the reliability of our Martian-sensitive cognitive faculties—but not before.

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 [wholly] 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 PD1].

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

Importantly, Thurow notes 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-based challenge—the *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 her 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 expands Jong's and Visala's two earlier points, which were (1) CSR can address only 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: (3) a sufficient natural explanation, such as CSR, would not only fail to disprove the *object* of that belief, but also the *existence of justifying evidence* (even if no one is currently aware of it).

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* 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 *do use*, rendering CSR even more irrelevant to the justificatory status of theism than even Jong and Visala had argued: If our HADD-based

BFF is off-track in even a "perniciously deceptive" way, it matters little since that 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, sustained by justificatory reasons J_p, J_t, J_i and J_{csr}, each of which is sufficient on its own to sustain S's belief B in G. In other words, justification for B is *overdetermined* for person S. Consequently, *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 there are other "legs" holding up this "table," as it were, any one of which on its own might justify belief—and it is one or more of those other ones that most believers *actually use* for justification in any case.

Consider the following analogy. Let us say that we are in a room in which there is a realistic 3D projection of an apple. Let us assume first that there *is* an apple in the room, though it is made completely hidden within the 3D projection of an apple. 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 what's more, he can *see* it. His justification is *overdetermined*: Either of these justificatory conditions would satisfy him that there is an actual apple present. We can show that he would still see the projected apple, even if we remove the real apple, since the projector would continue to project. This means his sight-based justification is indeed unreliable. However, his *other* justifying reason for belief was *smell* and it *is* reliable, as he would no longer smell the apple, once it was

removed. Until we knock out all the justificatory legs of his belief, we have not shown his belief to be unjustified. Only if he continued to believe based *solely* on sight, would he no longer be justified in believing that there is an apple.

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 consolidate and summarize these objections, which will help chart our path forward.

As we have seen, Plantinga argues that belief in God is *properly* basic, contra classical foundationalism. Indeed, his proposed inductive approach for identifying properly basic beliefs provides strong (though not completely impervious) shielding for such beliefs, 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 defines warrant truth-conditionally. 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 at least 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 specifically argue against atheist-CSR. In doing so, they raise the naturalist-question-begging objection against claims that HADD is "wrong" and "over-detecting" whenever it produces specifically *supernatural* beliefs. The also raise the related EAAN-based objection to the claim that our *religious* beliefs

are "unreliable," since such a claim ends up "throwing the baby out with the bathwater." In other words, we cannot argue that our CSR-based religious beliefs are unreliable without rendering *all* our beliefs unreliable.

The other powerful line of criticism of CSR-based atheistic arguments can be called the "explaining is not explaining away" objection. [21] This includes, roughly, two basic types of criticisms, which can be grouped under the following fallacy categories: the Genetic Fallacy, and the Single Cause Fallacy. [22] The Genetic Fallacy, as applied to atheist-CSR, was expressed in a variety of ways, from the simple, "CSR at most addresses how we come to believe in God, not whether He exists" to the more formal Bayesian 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 even a *proven* sufficient cause for belief, such as the natural one posited by CSR, cannot be assumed to be the only cause. There may well be other causes, which can separately originate *and* sustain belief. Other causes might come not only from a higher, even "ultimate" causal level, but also from other sufficient causes operating at the same level. In fact, it was argued that the justifications most believers *actually use*, are not addressed by CSR at all.

---

[21] I have adapted this concise phrase from (Jong, Explaining Religion (Away?))

[22] These categories are being used here as conceptual tools of convenience and should not be thought of as mutually exclusive or jointly exhaustive.

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.

CHAPTER 4

CRITICAL ASSESSMENTS OF THE OBJECTIONS

We have seen the foundational role that Plantinga's Reformed epistemology and EAAN have played in the current debate. A critical assessment of its key assumptions is therefore prerequisite to assessing the arguments that have been built upon them. We will begin with an examination of Plantinga's claim that belief in God is rationally justified because it is properly basic. We will then critically assess his concept of *warrant* upon which he attempts to shift the burden of proof onto the atheist.

Martin's Critique of "Reformed Foundationalism"

Recall Plantinga's discussion of the Great Pumpkin Objection, in which he argued that he had at least as much justification in rejecting that and accepting Christian belief as the atheist did in doing the converse. His argument was based on his 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*: that each community was responsible to its *own* starter set of examples. Nonetheless, Michael Martin is keen not to let this pass without pointing out the extremely 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 Voodoo practitioners to fairy believers—not just atheists. All that would be required is that their beliefs be basic *to their members* (Martin 272).

While it seems that Plantinga would not disagree[23], Martin's full exploration of the implications of Plantinga's position highlights the fact that it is a rather extreme form of rational *relativism*.

The issue is not the idea of basic beliefs vis-a-vis foundationalism, 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 *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, or beholding a flower) can and do trigger a wide range of *mutually exclusive* beliefs—including specifically atheistic ones (e.g., as a response to beholding natural evil, or even the apparently God-directed

[23] The best way to defend this assertion is to quote 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, Faith and Rationality 77). While he was comparing only those two communities, it does seem the floodgates of relativism across *all* communities are thereby opened. To see this, simply replace "Christian," "atheist," "Bertrand Russell" and "Madelyn Murray O'Hare" with the relevant counterparts from other communities.

evil deeds as recounted in the OT). It is precisely this lack of agreement among epistemological peers that seems to preclude belief in God from being considered properly basic (Martin 275).

Martin also points out that Plantinga's attack on classical foundationalism seems to discount the *contemporary* epistemological landscape, which is dominated by more modern versions of foundationalism that seem less clearly vulnerable to Plantinga's critique. 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. And, as we have already seen, there are non-foundationalist epistemologies, such as coherentism (among others) (Berker), which if correct, would further weaken the basis for Plantinga's argument.

Given these issues, it appears that Plantinga has failed to establish that belief in God is properly basic, which means it cannot be admitted into that privileged and evidence-resistant status. But even if Plantinga had been successful is establishing that belief in God is properly basic, how effective is his warrant-based approach to shifting the burden of proof?

Dawe's and Jong's Warrant Defeater

Recall that Plantinga claims it is the unbeliever that 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 an important clarification of "reliability."

Rather than attempting to *define* reliability, they come at it from the other direction, identifying ways to recognize instances of *un*reliability in both the uncorrelated and negatively correlated cases (Dawes and Jong 8). Freed of this positive definitional requirement, 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 point out, this circumstance-specific approach is consistent with the way in which we think of the reliability of our perceptual senses (Dawes and Jong 7).

This simple narrowing of the *context* of reliability also helps to address one part of the EAAN. Recall that the EAAN 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—about *anything*. However, this claim is undermined if we limit the *scope* of this unreliability claim.

Of course, it is not quite that simple. Recall that Barrett and Church had anticipated and dismissed this move as nothing more than special pleading, at least until some *reliability criteria is established* that clearly separates religious belief-forming conditions from non-religious ones. In other words, without such *positive* criteria we cannot say that a BFF such as HADD is typically reliable in nonreligious-belief forming

contexts, but unreliable otherwise, without question begging (Barrett and Church, Should CSR 8). But Dawes' and Jong's approach shows that such *positive* criteria are not necessary; one need only recognize relevant cases of *un*reliability. Just as one can, in some situations, recognize a wrong answer (or a non-answer) even without knowing the details of the right answer, so too one can identify particular instances of unreliability without a fully articulated general theory of reliability. [24] This certainly seems correct. I may, for example, not know how an airplane flies, yet still know that whatever the answer is, it will *not* be that the gravitational constant around the airplane is changing.

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), that belief 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

[24] 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, Faith and Rationality 75).

realistic holographic images of both whiteboards and whiteboard markers, without informing lecturers. . ..” Based on other factors, it is likely 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 seems incorrect. As they point out, she should now at least want some independent evidence to offset the belief-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 *basic* belief. What is the difference between that example and the current whiteboard example? The answer, I suggest, is that the former uses the accused’s memory-based basic belief in order to attack the *veracity* of the prosecutor’s belief. 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. (As we have already seen, a warranted belief could be false, and an unwarranted belief could be true.) 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 belief (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. In other words, one is making an *inference to the best explanation,* an approach that avoids the need to import any atheistic metaphysical assumptions (Dawes and Jong 15). Moreover, because theirs is specifically a *warrant*-defeating argument, the Christian cannot appeal to their beliefs about warrant without committing their own form of question begging (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 achieves a number of things: (1) it diffuses the charge of naturalist question-begging when CSR data is used to argue that religious BFFs are unreliable while other, non-religious ones are; (2) it answers Barrett's and Church's objection that treating religious contexts and non-religious contexts differently amounts to special pleading; and (3), it illuminates (through the notion of inference to the best explanation) one particular 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 below when we consider the implications of parsimony. Meanwhile, we will explore some surprising implications of Plantinga's own arguments, which, in their own ways, are as self-defeating as he claims naturalism is.

Bergmann's Reidian-Based Defense of Naturalism

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), which can be 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 do *not* have a defeater for R given N and E.

He does this by examining Plantinga's own critical analysis of an atheistic Argument from Evil. Recall that additional background knowledge can change the 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 expression by collapsing N, E, D and all other relevant knowledge into *k*, which gives us P(R/*k*).

This can also 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 propositionally determined low probability of her innocence, she is nonetheless *reasonable* in believing that she is innocent.

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 *seeming*, a non-propositional *basic* belief that trumps that propositional knowledge. Bergmann points out, however, that this very *same* argument can be used *against* Plantinga's EAAN. His insight is based on Reid's development of his famous "commonsense" epistemology, which so influenced Plantinga's own. Reid's thinking on non-inferential basic beliefs is that they are rooted in our natural commonsense (a view that does *not* depend on holding beliefs in the supernatural). This natural commonsense generates basic beliefs directly from experiences, like sense perception, but 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." It does this by generating certain emotions, such as those associated with *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 "seemings" to override the weight of propositional evidence against his belief, so the naturalist can invoke her own non-inferential knowledge to override

arguments like Plantinga's (Bergmann 625). As Michael Martin put it, "since what is self-evident is relative to persons, a classical foundationalist (CF) 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 270-271).

This point is related to a commonly overlooked confusion in Plantinga's argument, which is best illustrated by examining his own thought experiment regarding a "hypothetical species," 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, WCRL 329-330)

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.

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 example gives us a defeater

regarding belief in cognitive reliability; however—and this is key—it is a defeater for *our* belief in *their* cognitive reliability, not ours. 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* guilt, since we have access to relevant non-propositional knowledge.

Even with this inner access to non-propositional knowledge, is belief in R 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 R. 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, WCRL 180). Exploring these implications of this conflict, however, turns out to be quite important.

Bergmann observes that while one's non-propositional belief in God may override strong propositional evidence against that claim, it could still be defeated by one's acquiring *other* non-propositional beliefs, such as this: Basic belief in God is the product of an Evil Demon (Bergmann 625). His point, of course, is not to defend such alternative explanations, only to note that *believing* them would constitute a defeater.

Bergmann has used Plantinga's own Reidian-epistemology-based argument to undermine his conclusions. First, he used Plantinga's argument against the atheist Argument from Evil to undermine the key EAAN premise that the conditional probability of human cognitive reliability is low. Second, he used Plantinga's argument that belief in God is basic as well as his critique of CF to also show that the naturalist can believe in cognitive reliability as a *basic belief*.

It is important to keep in mind that Bergmann's critiques do not depend on actually embracing Reid's epistemology. However, by showing that this same epistemology can be used to *defeat* the EAAN, his critiques amounts to an effective *reductio ad absurdum* argument against it. This was does in much the same way that the EAAN itself assumed the truth of naturalism in order to undermine naturalism. 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.

Griffiths and Wilkins and the Cost of Truth

We saw earlier that a valid critique of some atheist-CSR 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. But this important way of thinking about multiple causes cuts both ways.

In Plantinga's EAAN there is different kind of levels-of-explanation confusion. According to Griffiths and Wilkins, he is inappropriately casting complementary

explanations from distinct levels as competing alternatives at the same level (Griffiths and Wilkins 5). Recall that Plantinga had powerfully argued that behavior and belief-content are very different things. Given that evolution can "reward and punish" only behavior, there seems to be no selective pressure on belief-content—that is, on *truth*-tracking—only on *fitness*-tracking. Griffith 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 can be broken down into at least two levels. One explanatory level answers the question, "Which *aspect* of some particular trait is the *target* of selection?" The other asks, "Which trait is more *fitness enhancing* relative to other traits, regardless of trait details?" 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 relates to the belief-content versus behavior distinction we need to understand the role of *cost* in evolution.

Arguably, Griffith's and Wilkin's most important contribution to this discussion is their reminding us of a central evolutionary concept that is 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 improvement. "Cost" can be energy, material expenditure, or even deleterious side effects. Importantly, it can also be *time*. For example, consider an investor (my example). 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 has all 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 would far outweigh the investment's best-case return). Let us say this describes the daily reality of all of her investment decisions.

In order to survive, she develops some *imperfect*, but much-better-than-chance *rules of thumb*. From a levels-of-explanation perspective, this partial retreat from perfect accuracy would not change that fact that she is *still* after the best investment choice. She is not now after something *else*. We might wonder why some particular rules of thumb work for her, especially since, in certain situations, they can and do result in bad investment decisions. Likely the *details* of her rules reflect contingent facts about her local resource and threat environment. Were they different, her *particular* rules would be likely be different, but that is because they would be adjusting to track *the same thing* as well as possible given the constraints.

This same notion of constrained optimization 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 perfect state, such as being able 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 links our cognitive faculties directly to evolutionary pressures, 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 intrinsic *logical* structure of many cognitive tasks. When forced to make a decision under conditions of uncertainty, the risks of false positive and false negative errors move in *opposite* directions. One can be reduced only at the cost of increasing the other. Evolution, therefore, would be expected to favor those who tend toward an optimal 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)—on average for the population.

The truth-tracking heuristics that would evolve under such cost-constrained conditions would, almost by definition, lead to cognitive *biases,* since the expected payoffs (i.e., net benefit [positive or negative] when correct versus when incorrect) are often strongly asymmetric. These biases, then, would reflect the kinds of “rules of thumb” that work most of the time (on average for the population) *in the situations in which they evolved*. 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” [italics added] (Vlerick and Broadbent 189).

Accordingly, Griffiths and Wilkins analyze the evolution of cognitive reliability as a straightforward instance of *decision making under uncertainty*. A creature has to "decide"[25] 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 (positive and negative) of each of the four possible combinations of action (act or do not act) and signal accuracy (veridical or not) (Griffiths and Wilkins 8).[26]

Griffith's 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 this actually addresses Plantinga's basic distinction between adaptive behavior and belief content. Plantinga himself recognizes that truth must be accurately tracked somewhere—the evolution of adaptive *behaviors* could hardly have occurred otherwise. The issue is whether *belief content* also tracks

---

[25] It cannot be stressed enough that such evolutionary analyses should never be construed as saying that creatures (even humans in many 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.

[26] 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.

truth. Griffiths and Wilkins had wisely assessed the payoffs in terms of the four possible combinations of *action* and *signal accuracy*, rather than *action* and *belief-content*; however, the strong evolutionary link between signal accuracy and belief content was not fully developed. Nonetheless, their insightful analysis can be directly built upon to make this strong connection. How? Once again, the answer is directly related to *cost*: "The human brain makes up about 2% of body mass, but accounts for about 20% of oxygen consumption. Having beliefs, whether true or false, comes at a high price" (Griffiths and Wilkins 6)[27]. True enough, but we have to be careful not to move too quickly from behaviors to "having beliefs," lest we beg the question.

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, WPF 225). He then looks at the range of possible independent desires and beliefs that could combine to produce behavior (Plantinga, WCB 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 2+2, 6+ (-2), etc. Given such an indefinite range

---

[27] Biological complexity of any kind, not just the brain's, has to *pay its way*, as it were (Dennett). That is, whatever survival benefit such complexity creates must exceed the cost not only of creating it, but also of maintaining it. As an extreme example of this, Dennett reminds us of the Sea Squirt, whose nervous system "pays for itself" only while the creature is looking for a permanent home. Once that home is found, however, the benefit of the structure no longer outweighs the cost of maintaining it (given its ecological niche). The result? The Sea Squire literally eats its own brain (Dennett 366-377).

of possible belief-desire input combinations that could produce a particular behavior, it certainly seems that the (Vast) majority of beliefs would be false just on statistical grounds. With our cost-based understanding of evolution, however, we can see 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' point that we are sometimes forced to make epistemic decisions on insufficient evidence is (perhaps ironically) exactly the point (James).

Plantinga's is a r*eductio* challenge, based on his assuming a naturalist framework in order to undermine naturalism. In that context, however, belief content cannot be seen as the free-floating *costless* mental phenomenon that dualists seem wont to imagine. Rather, from a naturalist perspective, it must be seen as *entirely* determined by physical brain activity. Accordingly, we have to ask what it would take for evolution to create a cognitive system that, in response to environmental cues, would produce, more often than not, a *false belief* (even in combination with some unspecified desire) such that it produces *adaptive behavior* that would be *selected for*.

Griffith and Wilkins argue that this would require a second cognitive mechanism, which would have to track truth in order to make the adjustments necessary to ensure the *behavior* remains adaptive while keeping the *belief* false (Griffiths and Wilkins 6-7). Given the assumed direct connection between the physical and the mental, such inefficient complexity would "cost" more. It would require more energy in much the same way that a more complex, inefficiently written computer algorithm would require

more electrical energy than a less complex, more efficient one. It would, in short, be very expensive relative to the simpler, more direct, *belief*-truth-tracking approach.

Their point is well taken, though it seems to address only the case of a *deceptive* cognitive faculty, rather than one that is simply *orthogonal* to truth. That is, this second cognitive mechanism would only be needed to keep the beliefs *false*, as if falsity were the objective. A more robust, cost-based, general case for linking belief content to behavior can be made from the following standpoints: 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 situations in which it would produce *behaviors* with negative payoffs. On the other hand, the less truth-tracking belief content is, the fewer the situations (possibly just one) in which it would produce high-payoff behaviors, and, of course, the greater the number of situations that would produce low and even negative-payoff behaviors. In other words, better *belief* truth-tracking would provide a higher behavioral "return on investment," where the "investment" is the energy-cost of creating and maintaining the necessary cognitive complexity associated with creating and sustaining that *belief*. (Note that adding desire to this produces the same result since desire links belief to behavior. The same forces would be expected to line these up with true belief in precisely the same way.)

Significantly, this concept is related to one of the attributes of a good scientific theory. Among other virtues (which we will review and justify 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 that they explain. Predictive success is an important aspect of this simplicity, since it avoids the need to encumber (read "add unnecessary complexity to") the original theory with ever more *ad hoc* hypotheses to "explain" phenomena that violate the original theory's predictions. In the extreme case, the proliferation of *ad hoc* hypotheses would simply cause the "theory" to devolve into a detailed *description of* the phenomena it is intended to "explain." It would end up reproducing the phenomena it in all its unexplained detail. It would, in other words, cease to be a theory at all.

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 whose desire it is 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, WPF 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 selection path whereby such a situation would gradually emerge (remember, each step has to confer advantage—or "pay for itself") and be maintained in a population. How, for example, should we imagine this individual responding to another large predator, one that she had never encountered before? 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 more situation-specific rules and therefore

far less parsimony, 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 is a more energy-intensive structure by virtue of its neurochemical complexity. It is also difficult to see how it would provide a *competitive advantage* over beliefs that happen to be more even a little more truth-tracking—that is, beliefs that align more (even a little) with ontological categories, like Predators, Persons, etc., and thereby enhance survival in a far wider range of *novel* situations.

Fales reaches a similar, but even more far-reaching, conclusion from the standpoints of deductive inference and enumerative induction (Fales 442-443). 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, such as our very ability to reason, impossible. Why? Because, once again, there would be no plausible step-wise natural selection path to get there. There would 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 external truth could such evolution-driving feedback occur. Note that this argument respects the distinction between belief content and behavior: in order for an *algorithm* to evolve, it is belief content in the form of true premises that would have to track truth.

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 here in order to, hopefully, make the point more robust. Freddy falsely 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. Later, however, he falls into the sea and, acting 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 remaining in the gene pool, since those *beliefs* would lead to more adaptive *behaviors* in more *diverse situations* (for the same cost—that is, the cost of maintaining the single neurochemical structure associated with that belief) than those holding beliefs that were less close to truth—that less accurately mirrored reality’s ontological categories. Once again we see that there are “no effective algorithms connecting false belief to *appropriate* action, as there are when the input is true beliefs and the rules of inference employed are valid or inductively sound” (Fales 443).

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 in very many different ways over many generations. A fortuitous combination of false beliefs might help in one or even a few interactions, but over time, across the population, and across disparate situations, the “cost” of having a

fundamentally wrong belief would quickly manifest itself as it is 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 entirely 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.

To the logical critiques of the EAAN that we had examined earlier, Griffiths and Wilkins add 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 show how cost-constraint considerations ensures that fitness-tracking would not only track as much "signal" accuracy as it could afford, but also *belief-content* accuracy. Fales bolstered this further by pointing out that our very ability to deduce, to *reason*, could never have evolved were our beliefs systematically false. We have to conclude from this analysis that belief content, no less than behavior, is a target of active, cost-constrained 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 does not mean our beliefs become justified by default. With the help of

Griffith, Wilkins, Fales, and others, we made some progress toward a positive case for the reliability of our cognitive faculties and the beliefs these faculties produce. Thus far, the case seems strongest for those faculties that are directly subject to selective pressures, for example, grasping enough of reality to find food and avoid predators. But this seems incomplete. Can we climb from these "commonsense beliefs" to our more rarified scientific and philosophical beliefs? It seems we can.

Vlerick's and Broadbent's Humean Bootstrapping

The EAAN argues that *none* of our beliefs is reliable (assuming atheistic, or, at least, *unpurposive* evolution). Therefore, if our preceding discussion has refuted it, we have succeeded in showing only that some of our beliefs are reliable, in particular, those that are clearly directly related to survival in our ancestral environments. Yet it seems that many of our beliefs would remain suspect. For example, why should we think our beliefs in advanced philosophical and scientific concepts are reliable, given that these domains did not even *exist* for most our evolutionary history? In other words, even if fitness-tracking *is* belief-truth-tracking, then 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. The commonsense truths that it tracks are often 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 our *commonsense* cognitive faculties—our "intuitive ontologies" (Vlerick and Broadbent 199) and cost-based rules of thumb—we still have a "bridging" problem. How do we build a justificatory bridge from our commonsense beliefs to our more abstract beliefs, like our scientific and philosophical beliefs? While some have argued for a kind of transfer-of-justification, such as the Milvian Bridge argument (Griffiths and Wilkins 3), this overlooks the fact that our commonsense faculties actively *obstruct* our efforts to grasp broader, deeper truths. The result is that many of these truths seem absurdly counterintuitive. Consider, for example, the wildly counterintuitive nature of the now extremely well-confirmed theory of quantum physics (e.g., an object'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's and Broadbent's "Humean bootstrapping" approach not only addresses this worry, but also suggests a more general strategy for addressing a key aspect of this paper's broader subject: CSR's implications for the rational justification of theism. Their insight is to recognize that the cognitive-reliability question should be approached from *within*, as it were, rather than as admitting of a single, general, outside-looking-in answer that would provide 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, in some sense, self-*correct*? Earlier we observed how our evolved commonsense intuitions often create obstacles to grasping more scientifically accurate, yet deeply counterintuitive, truths. But the fact remains that we evidently *have* overcome them, as the (self-evident) progress of science and technology would seem to show.

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" appears able to cut across many of the others, as if it overlays them. For example, our compulsive usage of analogical reasoning and metaphor would appear to be cases of our using cognitive skills from *one* specialty area to reason about *another*. This suggests that humans can *override* the deliverances of their individual modules' evolved intuitive ontologies in a way that other species cannot. The idea is that there is, apparently only in humans, 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" [cited in] (Vlerick and Broadbent 200). Importantly,

Vlerick's and Broadbent's argument does not depend on the truth of the Standard Model's modularity view of human cognition, or, indeed, of any particular evolutionary story of how our cognitive abilities arose. Rather, it illustrates an approach that echoes Dennett's description of bootstrapping in another 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" [italics 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 believe) 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 one moves to a new country and knows *no one at all*, how can one 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. (Indeed, this describes 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 an algorithm—not the *whole* of the solution. Through the buildup of experience (multiple iterations/steps) with this first randomly chosen group of people, one can gain or lose trust and from this accumulation of knowledge establish better-than-chance introductions (by increasingly trustworthy people) to other potential friends, who can replace those who were dropped. The process continues, iteration by iteration, with later iterations building upon the results of earlier, less-informed ones.

While this seems commonsensical, it is a rich analogy for our purposes. It can, for example, answer Plantinga's global skepticism worry. Recall that Plantinga argued it is self-refuting to believe, simultaneously, in naturalism, evolution, and the reliability of our cognitive faculties. His solution was to look *outside* of naturalism, to God's design, in order to avoid the bog of radical skepticism. But note that his approach assumes the need for a *single-step* solution to the question of how we can trust our cognitive faculties (i.e., God designed it). Vlerick and Broadbent, on the other hand, answer this question not with a single-step solution, but rather a multi-step, algorithmic, cumulative justificatory process. Indeed, it is precisely this foresight-less, 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 inputs and cognitive domains, would *itself* become a target of selection. If we add to this Fales' point about the necessity of true premises in the development of algorithms, we seem to have a very plausible, step-wise, selection path through which reliable, general reasoning could emerge—the kind of *general* reasoning that would justify our confidence in our ability to acquire even far-from-commonsense scientific and philosophical knowledge.

Summary

At this point we have come a considerable way in rebutting some of the key objections to the atheist-CSR position. To get a better sense of what force remains in those original objections, we will first summarize where the discussion now stands.

We have argued that Plantinga's case that theism can be properly basic fails. Independently 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 allows us to confer reliability onto our more abstract beliefs. This case was built on the

general algorithmic idea of “bootstrapping,” which, as we will see, has even broader applicability.

CHAPTER 5

BUILDING UPON THE CRITICAL ASSESSMENTS

While this recent literature has clearly helped us reduce the force of the original objections to atheist-CSR, important elements still stand. Some of that resilience comes from a number of more general issues that have yet to be dealt with. 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, including 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 a necessary element in establishing a relevant connection between CSR and the justification for theistic belief.

In addition to making the case that science is at least relevant to supernatural claims, we will need to separately address a particular type of 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 warrant-based attempt to shift the burden of proof, there may be other reasons to suppose that the burden lies with the atheist. As a result, we still need to make a *positive* case that the burden of proof does indeed lie with the theist.

These issues will be addressed in parallel with the development and justification of a closely related, mutually reinforcing set of concepts: (1) the bootstrapping concept as applied to certain normative epistemological principles; and crucially (2), some key 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 application of certain methodological principles and analytical tools for assessing the impact of observational evidence on the rational justification of religious belief generally. We will then make just such an impact assessment relative to CSR's findings specifically. 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 reasons people *actually use* to justify their belief.

NOMA

The topic of this paper—and any other paper that attempts to use scientific findings or methodologies to evaluate 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 this paper, 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" [cited 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).

Fishman provides a particularly helpful Bayesian framework, with 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 816-817). While we will defend its relevance to *super*natural claims in a later section, for now we will simply point to the fact that Bayes is used by Plantinga[28] and other critics of the atheist-CSR position when making their own cases. A brief recap of Bayes Theorem is, therefore, in order.[29]

Bayes Theorem Explained

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

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

[28] See, for example (Plantinga, WCB 230)

[29] 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 a familiarity with Bayes Theorem. In the interest of clarity and brevity, therefore, it seems best to present the overview now with the God hypothesis included for illustrative purposes.

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 *prior* probability is expressed as P(H), which, as we saw earlier, is based on all the other background information we have, *excluding* the hypothesis 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.

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, this may mean little until we consider the effects of the other "negation" terms in our Bayesian expression.

The negation of the hypothesis, ~H, is worth a brief clarifying comment. If P(H) is the prior *probability* of the hypothesis (some value between zero and one), then the prior probability that H is *false*, ~H, would be expressed as P(~H). But this would just be one minus the prior probability that H is true. In other words, P(~H) = 1 - P(H).

P(E|~H), then, is the probability that the new evidence would be found if H were *false*. (Or, if you prefer, "given that H is false.") The importance of the P(E|~H) term becomes clear with a simple example: Suppose my hypothesis is that little green fairies are the sole, direct cause of all astronomical phenomena. This H, therefore, entails that the sun will come up tomorrow, since, *ex hypothesi*, fairies cause this. Of course, observing the sunrise tomorrow should not be much cause for excitement among fairy believers because we fully expected this E for *other reasons*. So, while P(E|H) is high, so is P(E|~H). Since P(E|~H) is in the denominator of our Bayes expression, a larger value puts downward pressure on the whole right-hand side of the expression, and therefore, on the left-hand side as well: our final result, P(H|E). In other words, a higher P(E|~H) weakens E's contribution to the likelihood that H is true.

Fishman points to the real-world example of experimental control groups, such as those given placebos as part of determining the effectiveness of certain drugs. If everyone who takes the *actual* drug reports a benefit, celebrations should be put on hold until we see what the placebo-taking *control group* reports. Think of the "I feel better" report as the evidence, E. If we get this evidence from the *control group*, it increases P(E|~H), which puts downward pressure on the posterior probability of H (that is, the likelihood that the drug is effective) *given* the evidence, E.

These "~H" terms, Fishman argues, are particularly relevant to both the scientific evaluation of, 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 supposedly supernatural phenomena. In particular, he assumes that

a natural explanation can be counted *as part of ~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 appear to be an assumption in need of defense. To begin this defense, and to begin laying the groundwork for the main argument of this paper, we need to look a bit more closely at the notion of *prior probabilities* (also referred to simply as "priors").

To bolster our intuition of *priors*, imagine a complete stranger telling you he has 3, 20-dollar bills in his pocket. What should be our attitude toward this claim? The claim is not only logically *possible*, it is consistent with background knowledge that suggests his claim would be an unsurprising scenario, with no "going-in" reason to suppose that it is more likely than not. Since this is a question of *prior* probability—that is, *without* any claim-specific evidence, but *with* all our background knowledge—it seems that "agnosticism" would be the appropriate attitude. What if, instead, this stranger claimed to be an extraterrestrial with a thousand times the physical strength of a typical human? Like the first claim, this one is also logically *possible* (i.e., it is not incoherent, in which case the prior would just be "0"). However, limited only to our background knowledge, it seems we should assess the prior probability of this claim as *much* lower than that of the first claim, despite the fact that both "hypotheses" are possible and that we lack relevant evidence in both cases. To be sure, saying the prior for the alien-hypothesis, or even for God, is *low* is not the same thing as saying it is *false*. There may be other factors that more than compensate for this low prior, and lead to a *high* posterior probability. These intuitions will be developed more formally below. First, however, we need to address

one critically important general point, which relates not only to Fishman's assumption that natural explanations are exclusive of God-based explanations, but also to the central argument of this paper.

Explanatory Virtues and Hypothesis Bundling

One can always posit some additional claim to make God's existence consistent with a natural explanation. As we have already seen, with the contribution of such "helper" hypotheses, God can be understood as an ultimate explanation, which *complements* the newly discovered natural explanation.[30] Implicit too in such an approach is the assumption that the burden of proof rests with the skeptic to show that such an interpretation is *not* the case. In other words, the theist is saying, "Since God can be made [however imaginatively] *logically consistent* with your explanation, it fails to *dis*prove His existence or that belief in Him is unjustified." 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 to our discussion. Consider: Do we believe *everything* that we cannot *dis*prove? Clearly, we do not. As we will see, the old adage is often true, but not always. Our Bayesian framework can once again shed some light. Recall our fairies "theory," which includes the prediction that the sun will come up tomorrow. The

---

[30] 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 more (or less) ancient times, God (and other gods) was very much considered the *proximate* cause of most unexplained phenomena.

problem, as we saw, is that this "evidence" is no less expected if that theory is *false*. If, however, our theory predicts (or entails) something that we would *not* expect to find, then finding it would indeed bolster (confirm) the theory. This 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. Based on our preceding discussion, we can now understand this in Bayesian terms, that is, in terms of how high the probability of E is given ~H. [31]

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., diseases, earthquakes, etc.) are unexpected under the omnibenevolent (Abrahamic) God hypothesis, H_G. Finding such evil is, according such thinkers, disconfirmatory evidence for the existence of an omnibenevolent God. There is a rich literature on this Argument from Evil, but 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 God Hypothesis, H_G, *on its own*, may predict the absence of natural evil, H_G + A, where A is some auxiliary hypothesis, does *not*. For example, if 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, even *ad hoc*, hypothesis seems to present a general problem, one famously captured in the Duhem-Quine Thesis. Roughly, this

---

[31] Of course, it not nearly so simple. There is much more that needs to be said regarding *falsification* and how one can seemingly always "save" a theory by adding an auxiliary hypothesis. We will explore this important point shortly, but for now, we will keep it simple for the purposes of illustration.

thesis can be understood as saying that unless one can *prove* that no theory-saving auxiliary hypothesis exists for a given theory, then any unexpected evidence, like falsified predictions, can *never* refute it. Some A can always be invented to save it. Kitcher describes philosophers' efforts to develop falsifiability criteria without accounting for Duhem-Quine as having resulted in the commonly-held view he calls "naïve falsificationism" (Kitcher 42).

As he points out, however, no theory is tested *in isolation*. For example, even simply looking through a microscope to confirm the presence of a particular microorganism, in a particular sample, "bakes in" assumptions about 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 collections of hypotheses, and all of which, taken together, have led us to expect to find this microorganism in this sample, and to interpret what we find in the way that we do. This interpretation might be that the observed microorganism confirms some new hypothesis about its ability to survive in certain environments. But seeing it under the microscope supports this hypothesis only if *all* the other hypotheses bundled into this test are *assumed* to be true.

How skeptical should this make us? Not very. Based on the self-evident progress of science and technology, we do seem generally able to separate the "wheat from chaff" with regard to good and bad theories. Our methods apparently work well enough, even if we are not always clear why. To help us get some insight into why they work, Kitcher reminds us of Duhem's own succinct characterization of theory confirmation:

"Hypotheses are tested in bundles" [cited in] (Kitcher 44). Kitcher's critical insight is twofold: "[1] While hypotheses are always tested in bundles, they can be tested in *different* [large] 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" [italics in original] (Kitcher 46).

The significance of these two points cannot be overstated. Consider the long and ongoing story of human knowledge accumulation *as an integrated whole*. From this perspective it is clear that we test not just in different bundles, but in *ever recombining* bundles, over and over again, in many different, shifting contexts. This means they share different *pieces* of *each other's* bundles. There is, arguably, a kind of *bootstrapping* effect here: As these activities continue, they mutually reinforce, iteratively improving the confidence of *each* bundled hypothesis by increasing the number of *different* successfully tested bundles of which it is a member. The probability that the bundled (embedded) hypothesis would be wrong in so many *different* (tested) bundles shrinks as the number and variety of bundles increases. The effect amounts to converging lines of independent corroboration. The result: Over time we create an ever-wider and ever-deeper web of ever-more confirmed hypotheses.

His second point, regarding the independent testability of auxiliary hypotheses, is equally crucial. Kitcher 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*), which together create a problem-solving strategy that can *predict* a wide range of often-unexpected phenomena given specific

observational inputs. Based on the observational data known at the time, the orbit of Uranus was predicted to follow a certain pattern. 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 accounting for this unexpected result. In the Uranus case, the auxiliary hypothesis added was that an as-yet undiscovered planet existed, and that it had the properties necessary to make celestial mechanics consistent with the observed behavior of Uranus. Why should such an auxiliary hypothesis be taken seriously? First, the new-planet auxiliary hypothesis can be tested *independently* of the theory itself. Confirming its existence could be done without assuming the truth of celestial mechanics—one simply looks through a telescope.

One could also have picked a different auxiliary hypothesis, such as adding a Uranus-specific equation to the theory. Why would this not be just as acceptable? Because doing so would encumber the originally 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.[32] In other words, it maximizes the "bang for the buck."

---

[32] In this particular example, nothing was added to the theory at all. Rather, the independent confirmation of the hypothesized planet just added to the theory's record of predictive success and, therefore, its probability of being a correct theory. Other situations might justifiably require an actual

Good Theory Virtues

This Neptune example provides some key initial insights into how we might separate good theories from not-so-good theories. As 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, Paglieri and Pigliucci). 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 (Kitcher), 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., one equation for Neptune, and another unrelated one for the next anomaly, etc.). The other is *independent testability*. It must be *both* testable, *and* testable independently of the problem being addressed. As an example of this last point, consider a theory to the effect that a little green fairy exists and is perfectly visible, but only if the fairy's equally powerful fairy adversary does not make her invisible. While there is an observational effect that can be looked for, it cannot be tested independently of the theory; as a result, the claim is not really *testable* at all. Good theories should also be *fecund*: they should lead to new areas of inquiry, which are themselves confirmed. And, as we have already

---

addition to the theory, though its value would, under this parsimony-optimizing guideline, have to "pay for itself" by explaining *more* than just the one aberrant data point. It would also have to integrate with the *existing* hypotheses, rather than just being some "orange tossed in with the apples" as it were. Recall that taking this disregard of *ad hoc* complexity to the extreme dilutes a theory into a *description.*

discussed, good theories should exhibit *predictive success*, in the sense of making *unexpected*, confirmed predictions, which is related to its being falsifiable in our "non-naïve" sense.

Justification

Why should we accept such criteria? Well, they are not being proposed arbitrarily. Again, the justification is rooted in our procedural *bootstrapping* approach. Whatever initial, even haphazard guesses we took (for whatever reason), they have been (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. As with all inductive knowledge, such conclusions are, in principle, *defeasible*, but this is hardly a skeptical result since any alternative methodology would have to be *at least* as effective as what it replaces. And, since many independent lines of corroboration converge on the *same* principles, the odds of a fundamental change, while never zero, can 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 intimately related ones, which are of particular relevance to our CSR discussion: Parsimony and Burden of Proof.

Parsimony

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, any less and the theory would 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 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 different propositions is 90%, then the probability that *all* three are true is 90% multiplied by itself 3 times, or ~73%). Therefore, the more postulated entity types there are in a theory, the lower that theory's prior probability, and, all *else being equal*, the lower its posterior probability. In other words, the more entity types there are, *the more ways there are to be wrong*.

Burden of Proof

This "all else being equal" qualifier leads us directly to a way of using Bayes to better understand Burden of Proof. The "all else" just mentioned includes new, relevant evidence. The *lower* the prior probability, the *stronger* such evidence needs to be in order to drive up the posterior probability. This is due to the fact that, as Pigliucci et al. point out, when such originally unexpected evidence is found, and to a degree sufficient to consider the theory to be well-established, then that theory becomes *part of the background knowledge* used in determining the priors in *other* contexts (Pigliucci and

Boudry). For example, the Germ Theory of Disease is considered well-established. If a new theory comes along that entails the Germ Theory of Disease is actually false, then the evidence for this new theory will need to be high enough to offset its very low prior probability. 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 ~H skeptic. Note that under this pragmatic approach, the burden can shift over time as evolving evidence affects background knowledge. 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 cleanly determine burden of proof in all cases. Indeed, the idea that one can determine formal rules *generally* for crisply identifying logical fallacies in real-world situations seems to run up against a "Fallacy Fork" (Boudry, Paglieri and Pigliucci): The more the rules reflect *realistic* situations, the less general their applicability, and *vice versa*.

The point is well taken. 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-unearthed WWII bomb *is* dangerous unless proven otherwise. Yet in a courtroom, we assume that the criminal *is not* guilty unless proven

otherwise. The difference is clearly prudential, or, alternatively, *cost*-based. Such considerations not only preclude precise, one-size-fits-all criteria; they ensure that whatever criteria is developed will depend on evolving, situational specifics. Yet recognizing these realities does not mean that clear cases do not exist, or that strong cases one way or the other cannot be made when those practical considerations are carefully accounted for.

Is Science Relevant to the Supernatural?

Now that we have reviewed some principles underlying Bayes Theorem and laid the conceptual groundwork for justifying and using certain epistemic 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 *super*natural claims. First, however, we need to show that these natural-world-based methodologies do not "stack the deck" against the *super*natural—that they do not *presuppose* naturalism.

The epistemological principles we have reviewed above have developed from our mundane natural-world experiences. Even Plantinga, for example, draws analogies between sensing the Divine and sensing the everyday, such as our sense-perception of a tree (Plantinga, Is Belief in God Properly Basic? 44). But this should not be taken to mean that our epistemological principles force us to believe only in what we can see or directly detect. 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 is because that while we could not directly (aided or

unaided) actually *see* the atom, its existence, given certain hypothesized properties, has *observational consequences*. Do supernatural phenomena or entities have observational consequences? Some certainly do. Consider claims such as telekinesis and ESP, often the object of experiments. Fishman, for example, reminds us of recent tests of so-called "psi" powers (Fishman 824).

We can readily imagine results that would have lent strong support to such claims. Similarly, we can imagine fairly unambiguous evidence of a *sensus divinitatus*, such as everyone's being born with a clear disposition toward to, say, Christian beliefs—even those born in cultures that had never heard of Christianity. 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 reviewed above. As Fishman puts it, the original question of how to epistemologically *demarcate* between "natural" and "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, whether supernatural or not (Fishman 830).

Nonetheless, there are prerequisites that such claims have to meet before they are confirmable *in practice*. As we have seen, it is possible to evaluate claims of invisible, even highly exotic, entities and processes (whether natural or supernatural), but *only if* they have *observable consequences*; and further, these consequences (or predictions) must be testable independently of the "theory" that entails their existence. But there is a crucial caveat that has not yet been pointed out: these observational consequences must be *publically available*—that is, they must be *intersubjectively verifiable*.

Does God have *public* observational consequences? Thinkers such as Plantinga believe that our brains embody something like Calvin's *sensus divinitatus,*[33] which clearly produces publically observable *effects*, such as our measurable predisposition to believe in supernatural agency. Indeed, this *sensus* amounts to a kind of extrasensory perception, which produces an experience that can lead to the kind of theistic belief that Plantinga considers *properly basic* precisely *because* it is the result of the supposed proper functioning of this very mechanism. This seems to place the basis for Plantinga's theism-as-basic-belief claim firmly within the broader context of *religious experience*. After all, the supposed existence of a *sensus divinitatus* naturally leads us to wonder just what the subjective deliverances of this "God-sense" are like. What kind of *experience* is this?

Michael Martin summarizes Swinburne's helpful religious experience classification scheme (Martin 155-156): Type I is the experience of a public object as a supernatural being, for example, seeing a snake as a demon. "Public" here is intended to mean that the presence of the object is intersubjectively verifiable under ordinary conditions, although not everyone would experience it *as* a supernatural entity. Type II is the experience of a supernatural entity in its supernatural form, e.g., the devil with horns and pitchfork, an angel with halo and wings, etc. Note that this is also intersubjectively verifiable: anyone else near the observer would have observed (in the ordinary sense) the

---

[33] To be clear, Plantinga does not reject evolution, just the idea that it is *unguided* (Plantinga, WCRL 129), which is why he can think of the *sensus divinitatus* as a product of physical evolution.

same thing. Type III is the same as Type II but without this intersubjective verifiability: the object would not have been seen by anyone else, even by someone standing right next to the subject. Types I through III are easily *describable* in ordinary language. Type IV categorizes those experiences that are described either by saying what the experience is *not*, or by using "paradoxical" language, or by simply saying it is indescribable. Despite the differences between them, types I-IV share the feature that the experience involves some kind of *sensation*.[34]

While any number of religious claims are of types I and II—that is, of *public objects*—the experience of God, which Plantinga defends as the basis for justified theistic belief, is of the non-public type. This is clear from his explicit *rejection* of a key classical foundationalist premise, which we reviewed earlier, namely, P is properly basic *only if* proposition P is self-evident or incorrigible or evident to the senses. Despite the fact that the deliverances of this supposed *sensus divinitatus* are experiences of a non-public object, the epistemic methodologies we have discussed still apply *provided* that this non-public object produces *public* observational consequences. And, as we discussed, the theism embraced by mainstream Judeo-Christianity, and by Plantinga, clearly entails publically available observational consequences. We have to conclude not only that supernatural entities and phenomena are within the reach of the methodologies we have been discussing, but also that the Judeo-Christian God is specifically so confirmable.

---

[34] Curiously, there is a fifth category, which contains "experiences" that involve no sensations at all, just one's being *conscious* of the supernatural entity. It is, however, completely unclear in what sense "being conscious of" something is not an experience. It makes sense to say, for example, "The experience of being aware. . ."

Understanding the deliverances of the *sensus divinitatus* as a special case of religious experience will allow us to examine Plantinga's argument from a different perspective; but first, we will pull together the concepts we have reviewed up to this point in order to firmly establish one of our key assumptions.

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. In other words, he assumes that a natural explanation is a member of "any mutually exclusive set of alternative hypotheses" (Fishman 816). Now, that we have laid sufficient groundwork, we can begin to address this. Let us being 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 is explanatorily superfluous relative to H_N.

This generalization entails other notions, such as inference to the best explanation. For example, if two theories explain/predict the *same data set*, but one explains more 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, *successful* theory. From a Bayesian standpoint, the prior probability of the more

(needlessly) complex theory would, therefore, be lower than the simpler one. Recall that this can be understood by thinking of extra elements (ontological types) as creating more ways of being wrong, with no additional explanatory power to offset ("pay for") that lower prior probability. This shows how, under some conditions, explaining really is *explaining away*. From this that 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 we had earlier identified.

CSR and the God Hypothesis

By way of recap, we have addressed the NOMA concern, and diffused the skeptical worry that comes from the Duhem-Quine thesis. We still need to make a positive case that the burden of proof rests with those interpreting CSR data as consistent with theism. We also still need to respond to the objection that CSR fails to address the reasons people actually use to justify their belief. Finally, we need to show that the findings of CSR have indeed further and significantly reduced the rational justification for theism, which we will do by applying the methodologies that we justified and introduced earlier.

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 pragmatic, situation-specific 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 an arbitrary exercise. 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 is with 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 having 3, 20-dollar bills and the other of 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 *exclusive* of the God hypothesis, H_G, the prior probability of H_G, $P(H_G)$, is *very* low relative to that of a natural explanation. Conversely, a natural explanation, given our background knowledge and exclusive of H_G, has a much higher *prior* probability. If this is correct, then the evidential bar for H_G is much higher than that for an alternate, natural, $\sim H_G$ hypothesis. The higher the evidentiary bar, the more the absence of evidence seems like evidence of absence after all. Fishman likens this to our reaction to the absence of evidence for Santa Claus, whose powers are far outside anything our background knowledge would lead us to expect. In the face of missing evidence for this red-clad entity, our natural response is not normally one of *agnosticism* (Fishman 817). More to the point, if we do find evidence for an alternative explanation for the apparent wide-spread belief in Santa, which is consistent with our background knowledge, then it does seem appropriate to treat it as an alternative, *competing* explanation—not as somehow *complementing* Santa.

While this analogy is insightful, one should always be particularly cautious of attempting analogies with God. To adapt this analogy to more closely track our developing CSR discussion, consider those who believe in Santa based on experiencing the spirit of Santa as if in a vision. Imagine that a class of drug is found, which in some doses induces “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 yet another alternative *competing* natural explanation for wide-spread belief in Santa, in the first case, and for God in the second.

Jeff Jordan anticipates one 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. Jordan is vividly illustrating the Single Cause Fallacy we examined earlier. His point? Just as our faculty for perceiving cups may be led astray, so too the *sensus divinitatus*. His response to this objection is particularly insightful: 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). (We might add to this our earlier observation that physical causes are also public—that is, intersubjectively verifiable.) And, as we have seen, this “otherness” is what reduces its prior probability, forcing it to bear the burden of proof as it raises the evidentiary bar.

God, it seems, is far more “other” than either Santa or our alleged space visitor. To be sure, that fact does nothing to *disprove* God (or the space visitor), but the wide

divergence of these claims from our *background knowledge* does raise the evidentiary bar, and, in clear cases (i.e., those with *very* low priors), raises it to such an extent that the burden of proof is clearly on the party making the unusual claim. With the burden clearly on those asserting God's existence, *some* objections to atheist-CSR are made irrelevant, in particular, those that simply point out that CSR's findings do not *dis*prove 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 *justification* for belief in it, let alone agnosticism toward it, and for analogous reasons.

But what about the levels-of-explanation objection? In this case, God is being posited as the *ultimate* explanation, and natural explanations, like CSR, are simply the proximate means through which He implements His Plan. It should now be clear why this is vulnerable to precisely the same burden of proof argument just made. However, even if the placement of burden of proof were still in doubt, we can also make the argument in terms of *parsimony* and *predictive success*: Does adding this additional ontological category (God) pay for itself? If we can explain the same range of the data without it, then it clearly does not. This leads us to ask how the alternative theories compare.

Parsimony and Predictive Success

Even though the burden of proof typically rests with the theist, and even though theism is about belief in a non-public object, the burden can in principle be met. For example, if a theory *entailing* God's existence makes unexpected predictions that are confirmed, this would strengthen that hypothesis—possibly enormously. Building on our

earlier discussion of comparative predictive success, we will first critically compare the *sensus divinitatus* with sense perception, and then apply a similar analysis to theistic and atheistic interpretations of CSR's findings.

Sense Perception and Religious Experience

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 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., the tree outside the window, other minds, memory, etc.).

Plantinga is far from alone in making such comparisons. Indeed, comparing the experience of God to ordinary sense perception is commonly made in casual encounters with believers. One can schematize and superimpose both types of experience as follows [adapted from] (Martin 157), where S is a spontaneous perceptual belief of a certain appropriate kind based on sense experience, and R is a religious belief of a certain appropriate kind based on a religious experience. C represents the conditions under which these occur, where C_E is the appropriate condition for a sensory perception, and C_R for a religious one.

(1) / (1′) S / R under appropriate condition C_E / C_R is likely to be true

(2) / (2′) C_E / C_R obtains

(3) / (3′) My S / R in (the tree outside my window) / (God) is of the appropriate kind.

(4) / (4′) Therefore, my S / 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 some reality external to, and independent of, the person having the experience—call this the External Causality Hypothesis H (Martin 158). H for Sense Experience we will call H_E and for religious experience H_R. In (1), the specific external trigger for the H_E experience is *intersubjectively* verifiable for each instance of sense perception. The same is true of the conditions, C_E. We expect that others under the same conditions, C_E, would have the same experience. What might a *competing* $\sim H_E$ look like for (1), that is, a *Non*-External Causality Hypothesis for sense perception? It seems that a *usefully* plausible one (i.e., something other than a Cartesian Brain in a Vat) would be difficult to construct—again, given the intersubjectively verifiable evidence that supports H_E and C_E.

On the other hand, a competing $\sim H_R$ (a non-external cause for religious experience) for (1′) seems all too easy to imagine, especially since we have extensive experience with a range of ~H causes for reported experiences of non-pubic objects—from drug effects and surgical stimulation, to various mental and physical disturbances. In the absence of evidence of some kind of pathology, we have historically had some, at least plausible, alternative psychological hypotheses, such as Freud's wish-fulfillment (Leech and Visala). Bear in mind that such alternative $\sim H_R$ hypotheses need not be flawless, only palpably *better* (based on our earlier criteria) than H_R given our background knowledge and that knowledge's effect on the priors. Moreover, the conditions C_R are similarly not public.

Significantly, things have changed since Freud's time. We have arguably never had a more plausible ~H_R hypothesis than since the emergence of CSR's *empirically based* findings. But are we not simply begging the question by saying these "extra sensory" belief-causes are false or unreliable? 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-predictive success criterion: We have good reason to believe these "extra sensory" belief-causes are unreliable because the resulting beliefs do not *independently corroborate*; instead, they paint no consistent or coherent picture. This is not question begging, but 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 they are compared with independent, converging lines of evidence that, unlike the H_R case, *do* support both H_E *and* ~H_R cases (such as CSR explanations). Again, such ~H_R explanations need not be flawless, only the best that can be inferred from the available data.

Theist vs Atheist CSR

In the specifically Judeo-Christian CSR interpretations we have examined (such as those appealing to the effects of the Fall), what does the entailed God Hypothesis, H_G, predict? Recall that such theism does have 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 findings of neuroscience, let alone CSR, would be *unexpected*.

Under a theistic interpretation we might expect that the essential "you" could be *muted* by physical changes (if we think of the soul as using the body as a kind of mouthpiece), but not essentially *changed*. However, evidence from brain injury, surgery, and other studies have provided consistent evidence that our personalities, identities, sense of self, feelings toward others, memories, etc., can be fundamentally changed or piece-meal extinguished by *changing the brain*.

The evidence is extensive and often shockingly counterintuitive. Beyond the obvious cases of short and long-term memory loss due to disease and trauma, the two most striking examples come from the effects of frontal lobe damage, and from the effects of *corpus calloscotomy* procedures—so-called "split brain" surgeries. 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).

The effects of the split-brain procedure are insightful in another way. 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 distinct selves*, each with its own "free will" operating quite independently of the other. When experimenters give instructions *only* to the right (non-talking) hemisphere (which can no longer internally communicate with left hemisphere) and it *acts* on those instructions, the unaware left hemisphere spontaneously confabulates a story to explain why it decided to act (Pinker 43).

The best inference from such evidence (given the considerations we have reviewed, such as parsimony) is that if such damage can *change* you, partially erase you, or remove the illusion of a single, undifferentiated mind, then the destruction of the *whole* brain would mean (again, most parsimoniously without conflicting with the data) the destruction of the *whole* essence of the individual. These neuroscientific findings are consistent with—indeed, often *predicted by*—an atheistic, unguided, evolutionary account of the mind's development (Fishman 822). On the other hand, these same findings are not only unexpected under a theistic (especially Christian) account, but require the laying on of *ad hoc*, untestable auxiliary hypotheses in order to maintain consistency with a version of the theory that is already encumbered by H_G.

However, in the specific context of CSR, there is 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. In fact, this is what Clark and Barrett suggest: "the results from cognitive science . . . suggest empirical confirmation for Thomas Reid's speculations" (Clark and Barrett, Reidian 650). What are its predictions and how do they compare with the atheistic interpretation of 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 to be the Abrahamic omni-God. The *sensus divinitatus,* then, would predict some global uniformity of belief that, at the very least, would not be *in conflict* with this 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. However, as we will see, even the weaker prediction is not met.

What we find is indeed 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). Yet the *universal* aspects of this phenomenon do not even vaguely approximate the omni-God of monotheism. 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 *universal* we can extract from this extreme variability is the common denominator of *agency*—the unchecked, flowering output of our inborn teleological "promiscuity" (Kelemen and Rosset), which is subsequently pruned back by our inborn cognitive architectures.

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. In fact, it is difficult to imagine *any* universal pattern that could not be accommodated by this particular approach. 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). Such auxiliary hypotheses, which are clearly not testable independently of the theory of which it is a part (Christian theism), seem to place theistic-CSR in the position of being able to "cherry pick" evidence: if it seems to confirm the theory, the evidence is accepted; if it does not, the Adamic Fall or the supposed intention of a supernatural deity is invoked. Of course, if a theory is consistent with any outcome, any turn of the data, then it is both unfalsifiable *and* devoid of explanatory power.

Yet there is a sense in which the Adamic Fall, in the context of a Christian CSR interpretation, *does* make a 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 diversification over time. Instead, what 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). To the extent that it is testable in the retrospective sense, theistic CSR is falsified.

Can a more parsimonious theory be invoked that meets our epistemological conditions (such as falsifiability) while explaining the *same* range of data that theistic CSR claims to explain? Recall that we had already addressed Plantinga's worry that unless our cognitive faculties reflected purposive design, we would slide into radical skepticism. With that worry set aside, it seems that simply *removing* the theistic element *creates* that more parsimonious theory—nothing needs to be added back. On the other hand, by leaving the God auxiliary hypothesis in the theory, yet *more* auxiliary hypotheses (i.e., the Fall, Satan's direct action, God's inscrutable intentions, etc.) must be added on an as-needed basis to ensure that the theory remains consistent with the data. The price of maintaining one unnecessary auxiliary hypothesis is to constantly introduce more. And, as we have seen, this not only adds complexity, it renders the whole theory unfalsifiable. However, if we simply remove the God auxiliary hypothesis, then the whole superstructure of other auxiliary hypotheses can simply be dropped. The theory not only regains parsimony, but also other key virtues, such as falsifiability.

As it turns out, unencumbered atheist-CSR not only explains the same data with much less, it better *predicts* what we actually find, since what we find is precisely what is expected if CSR's findings, such as HADD, are the products of *unguided* evolution. For example, recall that Plantinga had posited the *sensus divinitatus* to explain such things as our predisposition to *relational* beliefs involving God, which produce feelings of gratitude and guilt (Plantinga, Is Belief in God Properly Basic? 46). But this is already explained by atheist-CSR. Our social interaction inference engines produce rich, emotionally-intense, morally-laden, social exchange intuitions. Furthermore, it explains

not just guilt and gratitude, which would be expected to fire whenever a Person inference engine is engaged (recall that a god, being MCI, preserves this category and so retains such rich inferences), but also the *universal* and *particular* nature of many of our less-than-noble moral quirks. 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 correlate so strongly with *degree of relatedness* (Pinker 241-268).[35] Again, atheist- (or "simple-") CSR explains a fuller range of observed phenomena, without having to layer on untestable auxiliary hypotheses.

Summary

In order to make the case that supernatural claims are not in principle beyond the reach of natural inquiry and, further, that natural explanations compete with, rather than complement, supernatural explanations, we first developed an inductive case for a set of epistemological tools and methodologies. Using these we then made the case that not only are these tools and methodologies applicable to supernatural claims, but also that they do in fact show that natural and supernatural explanations are mutually exclusive.

Drawing on these methodologies, we made a positive case that the burden of proof lies with the theist when it comes to interpreting CSR data. However, considering

---

[35] See also (Pinker 269-282) 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 parents and their offspring.

the sometimes complex pragmatic issues that come into play when assigning burden of proof, we also reinforced our case by employing some of our other epistemic considerations, such as parsimony. Parsimony, as we have seen, is related to predictive success in the sense that any hypothesized explanation should be as simple as possible, but not so simple that it is wrong. These considerations were applied first to the general question of the reliability of any putative God-sensing cognitive faculty, such as the *sensus divinitatus*, and then to the specific subject of this paper: theistic versus atheistic interpretations of CSR data. In both cases we saw that the theistic hypotheses were either falsified outright or were encumbered with untestable *ad hoc* hypotheses in order to make them consistent with the observed data. In the latter case, we compared the encumbered theistic-based theories with their simpler, atheistic counterparts and saw that the additional theistic elements did not "pay their way."

It is through this same approach that we can finally respond to Thurow's charge that CSR is irrelevant because it does not address the reasons people *actually use*. First, to suggest that one needs to disprove each and every basis a person has for belief is to wrongly assume that the burden of proof is on the skeptic. But even if the burden cannot be clearly assigned, none of the other reasons people offer survive the test of predictive success. There is no corroboration for any putative supernatural sense perception in the way there is for ordinary sense perception as is apparent from the wide range of mutually exclusive "results." Moreover, if *ad hoc* explanations are brought in to explain this religious diversity, such as by invoking the effects of Original Sin, or the supposed intentions of a deity, they add complexity (and potentially push the original "theory" out

of the realm of falsifiability) while providing no explanatory or predictive power beyond what we get from the simpler, atheistic model--which *already predicts* what we find without the need for such auxiliary hypotheses.

It also is worth noting that the particular *ad hoc* hypotheses of Original Sin or the inscrutable intentions of a deity, are compatible with any degree of religious diversity imaginable. Perhaps here Barrett's and Church's charge of "special pleading" might be better applied to the theist. Moreover, as Dawes and Jong point out, when the Reformed epistemologist invokes Original Sin in this way, he 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). With Thurow's important objection addressed, we have to conclude that CSR, to the extent it is true, does indeed sharply reduce the rationality of theism.

In what initially seems like a surprising move Clark and Barret "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: belief in god is properly basic with respect to warrant, and so is not believed *as a hypothesis*. 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 these points fail. We have already shown why belief in god is *not* properly basic; and as for their critique of extreme parsimony, we have already seen that simplicity has a lower bound: a theory should be as simple as possible without being *wrong*. This collapse into the "black hole" of solipsism is possible only if one begins entertaining hypotheses such as the Brain in the Vat or "the universe was created ten minutes ago with our memories intact." These are not disprovable in the logical sense, but for reasons we have already discussed, there is no good reason to suppose the burden rests on those who disagree with such imaginings. Even more to the point, it is difficult to see how the God hypothesis is intended to shield against such logical possibilities. After all, perhaps God or an Evil Demon created *only* you *as* a brain in a vat. Perhaps as Bergmann earlier suggested, one takes it as a *basic belief* that an Evil Demon is responsible for creating false basic beliefs in God, which would be a defeater for that basic belief. This leaves us with Clark's and Barrett's original concession, namely, that there really is no reason to appeal to any god to explain the findings of CSR.

While the argument in this paper has not been limited to Christian theism, there is an additional potential problem specific to the Christian atheist-CSR objector. This problem 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 that if it is indeed God's Will that person S experience Him, then our finding *any* natural cause for God-experiences would provide good grounds for thinking that the discovered natural cause cannot be traced to God *at all* since the Divine

Will requires no such necessary natural preconditions (Jordan 262). This would further undermine any appeal to God as the ultimate cause operating through natural, proximate causes.

CHAPTER 6

ANSWERS TO OBJECTIONS AND PATH FORWARD

At least two objections can be developed using the methodologies discussed in this paper. The first objection is based on the pragmatics of assigning burden of proof. We had noted earlier that such considerations can shift burden of proof on a case-by-case basis, preventing the development of a simple one-size-fits-all rule. Pascal essentially makes this argument in his famous Wager. The second 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. Due to space limitations, these will be dealt with briefly. However, initial responses have been outlined here in the hope that they be developed further.

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 used.

Recall our discussion about the pragmatic *cost* considerations associated with assigning burden of proof. We used the example of an 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 Him. Does this not undermine our earlier argument that the burden of proof lies with the theist? There are at least three 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 contradictory—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 for not choosing them if they turn out to be right. So, the choice is not simply between *un*belief and belief, but also between many *beliefs*. After all, to be a believer in one system is to be an *un*believer 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: 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 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 that precludes theism on the grounds that it is not parsimonious, in the sense of not explaining the most with the least. Swinburne has argued in just these terms that, in fact, God *is* the simplest "explanation" of all (Swinburne). A full treatment of this is beyond the scope of the present paper; however, we can at least draw attention to the central problem: 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, and all our mental *states* (Ostrowick)—to be far more capable than our most *complex* computers and *complex* brains, and to hold emotional *states*, beliefs and desires. "God" as a *label* is certainly simple, but that label's referent is another matter altogether. But the problem with this "simplest" idea is deeper even than even this.

If we regard the notion of "explanation" as one of *increasing* intelligibility, of taking us from the *known* to the unknown, then God's radical otherness (such as being supremely simple while also exhibiting what is, in literally every other context, considered to be examples of extreme complexity) seems to preclude His serving as an "explanation" *of any kind*, let alone as a *simple* explanation.

Perhaps these concerns can be illustrated by way of an analogy. Suppose we are living during prehistoric times, before the natural processes behind 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: the tribal rain god. It is not only one simple word, but all conceivable rain-related phenomena can be understood in terms of this god’s metal and emotional states. In fact, she stipulates as matter of this god’s definition that it is *supremely* simple, even if we cannot grasp how that can make sense.

Again, I am only gesturing toward more fully-developed answers to these objections as part of suggesting a path forward based on this thesis.

CHAPTER 7

CONCLUSION

After introducing the cognitive science of religion (CSR), its historical development, and its current consensus as captured in the Standard Model, we illustrated what a CSR-based natural explanation for religious belief looks like. The objective was twofold: to illustrate the application of CSR concepts, and to give some indication of the strength of this field's evidential support, which is what lends urgency to the philosophical questions explored in this paper. Accordingly, we examined CSR's significance for the philosophy of religion—in particular, its epistemic consequences for the rational justification of theism. In the process, we saw that there is much wider agreement on CSR's findings and natural causal models than there is on the philosophical significance of the same. Indeed, given the very same *scientific* causal models, there are both theistic and atheistic interpretations, what we called atheist-CSR and its theistic objectors.

This paper is an attempt to persuade the reader that the theistic objections to atheist-CSR fail and that not only is the atheistic interpretation of CSR's findings the more rational, but also that these findings significantly reduce the rational justification for theism. To build this case, we began by examining the leading objector arguments. In the process, we surveyed the relevant state of the field in religious epistemology and discovered that Alvin Plantinga's important contributions to religious epistemology

generally, and to the present debate specifically, were wide-ranging and foundational, making the atheist-CSR objections serious indeed.

Plantinga's was seen as a multi-step approach that involves first, creating a space for rational, justified theism by undermining the rational *ground* of classical foundationalism and introducing theism as a properly basic belief (i.e., Reformed epistemology); second, shifting the burden of proof through his truth-conditional notion of warrant and his proper functionalist account of rationality; and finally, creating a rationality saving *need* for theism as an antidote to what he claims is the "corrosive" skepticism that results from an atheistic interpretation of the theory of evolution—this was his famous Evolutionary Argument against Naturalism or EAAN.

As we saw, others had built upon this foundational work in order to apply it specifically to the current CSR-inspired epistemological debate. In reviewing the current, leading objections to atheist-CSR we found a number of recurring themes: the naturalist question-begging objection against claims such as, "HADD is 'over-detecting' whenever it produces specifically *super*natural beliefs"; the EAAN-based objection to the common atheist claim that our *religious* beliefs are "unreliable"; the "explaining is not explaining away" objection, which is generally made by way of invoking the Genetic Fallacy and / or the Single Cause Fallacy; and finally, the relevance objection, the strongest of which was the charge that CSR fails to address the justifications people actually use. Cutting across all of these objections was the theistic burden-of-proof assumption, namely, CSR fails to meet its presumed burden to *dis*prove the existence of God.

To address these objections, we reviewed the historical and recent literature, and presented a multi-pronged response, arguing that (1) Plantinga's case for the proper basicality of theism fails, (2) that his notion of warrant fails to shift the burden of proof for God's existence onto the skeptic, (3) that the variants we encountered of the Question-Begging, Genetic, and Sole Cause fallacies can all be shown to be non-applicable, and (4) that the EAAN fails 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 "bootstrapping" case for the reliability of our reasoning abilities generally—not just our common sense beliefs.

In order to further build upon the current literature and, hopefully, to incrementally advance the debate, we needed to first build an inductive case for set of epistemological tools and methodologies, and justify them. Our bootstrapping-based justification for this set of methodologies allowed us to make the case that natural-based epistemological principles do indeed apply to supernatural claims and, further, that they provide some useful criteria for deciding when a supernatural explanation competes with, rather than complements, a natural one.

We also made a positive case that the burden of proof lies with the theist when it comes to interpreting CSR data. However, considering the sometimes complex pragmatic issues that come into play when assigning burden of proof, we reinforced our case by employing some of the epistemic considerations we had earlier developed and justified, such as parsimony. We then applied these considerations first to the general

question of the reliability of any putative God-sensing cognitive faculty, such as the *sensus divinitatus,* and then to the specific subject of this paper: theistic versus atheistic interpretations of CSR data. In both cases we saw that the theistic hypotheses were either falsified outright or were encumbered with untestable *ad hoc* hypotheses. In the latter case, we compared the encumbered theistic-based theories with their unencumbered atheistic counterparts and saw that the additional theistic elements did not "pay their way."

It was by application of this same general approach that we were able to address the charge that CSR is irrelevant to the reasons people actually use to justify their belief. Part of our response involved looking at this from the perspective of *predictive success* vis-à-vis religious diversity.

With that remaining obstacle addressed, we were ultimately able to conclude that the findings of CSR significantly reduce the rational justification of theistic belief. However, we identified two potentially serious objections could be mounted using the very same methodologies developed and advanced in this paper: one, a revitalized Pascal's Wager; and two, Swinburne's argument that God is the "simplest" explanation. Fuller responses to these can and should be developed. To begin that process, initial responses have been developed and presented in the Answers to Objections and Path Forward section.

WORKS CITED

WORKS CITED

Altschul, Jon. "Epistemic Entitlement." *Epistemic Entitlement*. (accessed January 14, 2016). <http://www.iep.utm.edu/ep-en/>.

Atran, Scott. *In Gods We Trust: The Evolutionary Landscape of Religion*. Oxford University Press, 2002.

Barrett, Justin L and Ian M Church. "Should CSR Give Atheists Epistemic Assurance? On Beer-Goggles, BFFs, and Skepticism Regarding Religious Beliefs." *The Monist* 96.3 (2013): 311-324.

Barrett, Justin L. "Cognitive Science of Religion: What is It and Why is It?" *Religion Compass* 1.6 (2007): 768-786.

—. "Exploring the Natural Foundations of Religion." *Trends in Cognitive Sciences* 4.1 (2000): 29-34.

—. "Is the Spell Really Broken? Bio-psychological Explanations of Religion and Theistic Belief." *Theology and Science* 5.1 (2007): 57-72.

Barrett, Justin L. *Why Would Anyone Believe in God?* Walnut Creek: AltaMira Press, 2004.

Bergmann, Michael. "Commonsense Naturalism." *Philosophy of Religion: an Anthology*. Ed. Pojman, Louis and Michael Rea. Wadsworth, 2012. 621-629.

Berker, Selim. "Coherentism via Graphs." *Philosophical Issues* 25.1 (2015): 322-352.

Bloom, Paul. "Religious Belief as an Evolutionary Accident." *Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion*. Ed. Murray, Michael and Jeffrey Schloss. Oxford University Press, 2009.

Boudry, Maarten, Fabio Paglieri and Massimo Pigliucci. "The Fake, the Flimsy, and the Fallacious: Demarcating Arguments in Real Life." *Argumentation* (2015): 1-26.

Boyer, Pascal. *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books, 2001.

Clark, Kelly James and Justin L Barrett. "Reformed Epistemology and the Cognitive Science of Religion." *Faith and Philosophy* 27.2 (2010): 174-189.

—. "Reidian Religious Epistemology and the Cognitive Science of Religion." *Journal of the American Academy of Religion* 79.3 (2011): 639-675.

Clifford, W.K. "The Ethics of Belief." *Philosophy and Contemporary Issues*. Ed. Burr, John R. and Milton Goldinger. Pearson Education, 2004.

Dawes, Gregory W and Jonathan Jong. "Defeating the Christian's Claim to Warrant." *Philo* 15.2 (2012): 127-144.

Dawkins, Richard. *The God Delusion*. Mariner Books, 2008.

—. *The Selfish Gene*. Oxford: Oxford University Press, 1989.

De Cruz, Helen and Johan De Smedt. "Reformed and Evolutionary Epistemology and the Noetic Effects of Sin." *International Journal for Philosophy of Religion* 74.1 (2013): 49-66.

Dennett, Daniel C. *Breaking the Spell: Religion As a Natural Phenomenon*. New York: Penguin Books, 2006.

Fales, Evan. "Plantinga's Case Against Naturalistic Epistemology." *Philosophy of Science* (1996): 432-451.

Fishman, Yonatan I. "Can Science Test Supernatural Worldviews?" *Science \& Education* 18.6-7 (2009): 813-837.

Gould, Stephan Jay. *Rocks of Ages: Science and Religion in the Fullness of Life*. Ballantine Pub. Group, 1999.

Griffiths, Paul E. and John S. Wilkins. "When Do Evolutionary Explanations of Belief Debunk Belief?" *Darwin in the 21st Century*. forthcoming.

Guthrie, Stewart. *Faces in the Clouds: A New Theory of Religion*. Oxford University Press, 1995. <http://0-search.ebscohost.com.torofind.csudh.edu/login.aspx?direct=true&db=nlebk&AN=151246&site=ehost-live&scope=site>.

Hauser, Marc D. *Moral Minds: The Nature of Right and Wrong*. New York: Harper Perennial, 2006.

Hume, Dave. *The Natural History of Religion*. London: A. and H. Bradlaugh Bonner, 1889. <http://oll.libertyfund.org/titles/340>.

Hume, David. "Against Miracles." *Philosophy of Religion: an Anthology*. Ed. Pojman, Louis and Michael Rea. Wadsworth, 2012. 406-415.

International, WIN-Gallop. "Global Index of Religiosity and Atheism." *Global Index of Religiosity and Atheism*. 2012 (accessed January 14, 2016). <http://www.wingia.com/web/files/news/14/file/14.pdf>.

Inwagen, Peter van. "Explaining Belief in the Supernatural: Some Thoughts on Paul Bloom's 'Religious Belief as an Evolutionary Accident' ." *Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion.* Ed. Murray, Michael and Jeffrey Schloss. Oxford University Press, 2009.

James, William. "The Will to Believe." *Philosophy of Religion: an Anthology*. Ed. Pojman, Louis and Michael Rea. Wadsworth, 2012. 502-511.

Jong, Jonathan and Aku Visala. "Evolutionary Debunking Arguments against Theism, Reconsidered." *International Journal for Philosophy of Religion* 76.3 (2014): 243-258.

Jong, Jonathan. "Explaining Religion (Away?): Theism and the Cognitivie Science of Religion." *Sophia* (2013): 521-533.

Jordan, Jeff. "Religious Experience and Naturalistic Explanations." *Philosophy of Religion: an Anthology*. Ed. Pojman, Louis and Michael Rea. Wadsworth, 2012. 258-266.

Kahane, Guy. "Evolutionary Debunking Arguments." *Nous* 45.1 (2011): 103-125.

Kelemen, Deborah and Evelyn Rosset. "The Human Function Compunction: Teleological explanation in adults ." *Cognition* 111.1 (2009): 138-143.

Kitcher, Philip. *Abusing Science: The Case Against Creationism*. Cambridge: The Massachusetts Institute of Technology, 1993.

Leech, David and Aku Visala. "Naturalistic Explanation for Religious Belief." *Philosophy Compass* 6.8 (2011): 552-563.

Lipka, Michael. "7 Facts about Atheists." *7 Facts about Atheists*. 2015 (accessed January 14, 2016). <http://www.pewresearch.org/fact-tank/2015/11/05/7-facts-about-atheists/>.

Martin, Michael. *Atheism: A Philosophical Justification*. Temple University Press, 1990.

Mithen, Steven. *The Prehistory of the Mind: The Cognitive Origins of Art and Science*. New York: Thames and Hudson, 1999.

Murray, Michael J. and Andrew Goldberg. "Evolutionary Accounts of Religion: Explaining and Explaining Away." *Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion*. Ed. Murray, Michael and Jeffrey Schloss. Oxford University Press, 2009.

Murray, Michael J. "Scientific Explanations of Religion and the Justification of Religious Belief." *Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion*. Ed. Murray, Michael and Jeffrey Schloss. Oxford University Press, 2009.

Ostrowick, John. "Is Theism a Simple, and Hence Probable, Explanation for The Universe?" *South African Journal of Philosophy* 31.2 (2012): 354-368.

Pascal, Blaise. "The Wager." *Philosophy of Religion: an Anthology*. Ed. Pojman, Louis and Michael Rea. Wadsworth, 2012. 496-497.

Pigliucci, Massimo and Maarten Boudry. "Prove it! The Burden of Proof Game in Science vs. Pseudoscience Disputes." *Philosophia* 42.2 (2014): 487-502.

Pinker, Steven. *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin Books, 2002.

Plantinga, Alvin and Nicholas Wolterstorff. "Faith and Rationality: Reason and Belief in God." (1983).

Plantinga, Alvin. "Is Belief in God Properly Basic?" *Nous* (1981): 41-51.

—. *Warrant and Proper Function*. Oxford University Press, 1993. <http://0-search.ebscohost.com.torofind.csudh.edu/login.aspx?direct=true&db=nlebk&AN=294767&site=ehost-live&scope=site>.

—. *Warranted Christian Belief*. Oxford University Press, 2000.

—. *Where the Conflict Really Lies: Science, Religion, \& Naturalism*. Kindle Edition. New York: Oxford University Press, 2011.

Raff, Charles. "Introspection and Incorrigibility." *Philosophy and Phenomenological Research* (1966): 69-73.

Ramachandran, Vilayanur S. "Beauty or Brains?" *Science* 305.5685 (2004): 779-780.

Smith, Martin. "The Epistemology of Religion." *Analysis* 74.1 (2014): 135-147.

Stirrat, Michael and R Elisabeth Cornwell. "Eminent Scientists Reject the Supernatural: a Survey of the Fellows of the Royal Society." *Evolution: Education and Outreach* 6.1 (2013): 33.

Swinburne, Richard. "God As the Simplest Explanation of the Universe." *European Journal for Philosophy of Religion* (2010): 1-24.

The Holy Bible: King James Version. Iowa Falls: World Bible Publishers, 2001.

Thurow, Joshua C. "Does Cognitive Science Show Belief in God to be Irrational? The Epistemic Consequences of the Cognitive Science of Religion." *International Journal for Philosophy of Religion* 74.1 (2013): 77-98.

Trigg, Roger and Justin L Barrett. *The Roots of Religion: Exploring the Cognitive Science of Religion*. Ashgate Publishing, Ltd., 2014.

Vegetti, Mario. "The Greeks and Their Gods." *The Greeks*. Ed. Jean-Pierre Vernant. Trans. Charles Lambert and Teresa Lavender Fagan. Chicago: Chicago University Press, 1995. 254-283.

Vlerick, Michael and Alex Broadbent. "Evolution and Epistemic Justification." *dialectica* 69.2 (2015): 185-203.

Wood, Graham. "Cognitive Science and Religious Belief." *Philosophy Compass* 6.10 (2011): 734-745.

Xenophanes. "Selections from Ancient Greek Philosophers." *Philosophy of Religion: an Anthology*. Ed. Pojman, Louis and Michael Rea. Wadsworth, 2012.