**The Goldberg Exaptation Model:**

Unifying Adaptationist and By-product Theories of Religion

Taylor Davis

Department of Philosophy

Purdue University

100 N. University St.

West Lafayette, IN 47907-2098

taylordavis@purdue.edu

**Abstract**

A central debate in the study of religious evolution has been that between adaptationist theories, which claim that religion is an adaptation, and by-product theories, which deny that religion is an adaptation. I argue here that this debate is empty. Because the concepts *religion* and *adaptation* do not have the same meanings in these two types of theories, the theories themselves are logically consistent. The question thus arises how they may be related, over and above mere consistency. I address this question here by presenting a model of the religious phenotype, the Goldberg Exaptation Model, which shows how adaptationist and by-product theories may be integrated. The model highlights the fact that the two approaches explain different functional features of religious phenotype, which have been acquired at different times in its evolutionary history. The Goldberg Exaptation Model thus clarifies the distinct role that each theory plays within a single, unified account of religious evolution.

**Keywords**

Religion, Adaptationism, Cultural Evolution, By-Product Theory

**1. Introduction**

A central debate in the scientific study of religion has been that between *adaptationists*, who claim that religion is an evolutionary adaptation (Sosis and Alcorta 2003; Sosis 2009; Haidt 2012; Powell and Clarke 2012; Bering and Johnson 2005; Irons 2001; Wilson 2002; Bulbulia 2009), and *by-product theorists*, who deny that religion is an adaptation (Boyer 2001, 2003; Pyysiäinen and Hauser 2010; Bloom 2007; Barrett 2004; Pinker 2006; McCauley 2004; Lawson and McCauley 1990). According to the by-product account, religion per se has no evolutionary function, and religious belief is explained as a by-product of the activities of cognitive traits adapted for other, non-religious functions. I argue here, however, that this debate is empty; adaptationist and by-product theories are mutually consistent. As a result, the question immediately arises how these theories may be related, over and above mere consistency. I address this question by identifying a model of the religious phenotype that has already been suggested, independently, by two theorists from opposing sides of the debate. Though the terms of the debate have obscured the commonalities, shared assumptions captured in what I call the *Goldberg Exaptation Model* are already sufficient for resolving the debate, and for unifying adaptationist and by-product theories of religion.

 The theorists involved take themselves to be in disagreement over the answer to an empirical question: is religion an adaptation? But while adaptationists answer “yes” and by-product theorists answer “no,” these two theories do not actually offer competing explanations for the same empirical facts. Instead, they offer different explanations for different facts, as a result of the fact that the concepts *religion* and *adaptation* do not have the same meanings. After first describing the positions in the debate, I argue that they are, in fact, mutually consistent. I then proceed to extract the Goldberg Exaptation Model from the existing literature, and use it to show how by-product theories and adaptationist theories may be integrated.

 The model was first suggested in metaphorical terms by philosopher Robert McCauley (2004), in defense of the by-product approach. He compares the religious phenotype to a Rube Goldberg device—a system composed of distinct sub-components that already possess their own, independent functions. However, a substantially similar model has also been proposed by an adaptationist theorist, anthropologist Richard Sosis (2009). In more literal terms, Sosis discusses the role of *exaptation* in explaining how the religious phenotype evolved, where exaptation is the process by which traits that already possess one adaptive function come to acquire new, additional functions. After first identifying the commitments that lie at the intersection of these two descriptions of the religious phenotype, I show how they render adaptationist accounts and by-product accounts not merely consistent, but importantly related. While by-product accounts explain how various component parts of the Goldberg device evolved in the first place, adaptationist accounts explain how these parts were later exapted to form a functionally integrated whole—the Goldberg device itself.

**2. By-product Theories of Religion**

While philosophers such as Hume (1992, 1993) and James (2004) have offered accounts of religion in naturalistic terms based on a materialist ontology, their theories have not stimulated actual programs of empirical research on religious belief and behavior. And while there is an ongoing tradition of empirical work on religion in social science (cf., Durkheim 1995; Weber 2002; Tylor 1871; Evans-Prichard 1956, 1965; Bellah 2011), these theories do not relate the social, phenomenological and philosophical phenomena they describe to theories of physical facts offered in the *non-*social sciences. However, in the last two decades researchers from across the disciplines of psychology, anthropology, economics and biology have begun to pursue the study of religion from an evolutionary perspective, and the adoption of this shared framework has led to the emergence of a new set of theories that are not merely consistent with what we already know about the living world in general, but that are capable of using this knowledge to develop richer and deeper explanations of religion. However, conceptual differences *within* this new field hinder continued progress toward this goal.

 The first of these two decades was dominated by an approach now commonly referred to as the *cognitive science of religion*, or CSR, and virtually all theories in this tradition are by-product theories.[[1]](#footnote-1) Adaptationists joined the discussion a decade later, but by the time of the first International Conference on the Evolution of Religion in 2007, Sosis (2009) reports that “One of the most heated topics of discussion concerned whether or not religion should be considered an adaptation or a byproduct.” He mentions this in an article devoted entirely to defending the adaptationist theory against a wide array of criticisms. Indeed, he explains that he is forced to limit his responses to only five of the most common objections. And he takes as an epigraph for this article a quote from psychologist Lee Kirkpatrick, a member of the opposing camp: “The first question to be addressed by any evolutionary approach to religion is whether religion is an adaptation or a byproduct of adaptations designed for other purposes.” So whether or not this is really *the* primary question in the field, it certainly represents *a* fundamental issue, one that turns, as I explain below, on the meanings of the fundamental concepts *adaptation* and *religion*.

 The CSR tradition focuses specifically on cognitive traits involved in forming beliefs about supernatural agents. It also focuses specifically on traits that are more or less universal among humans (cf. Boyer 2001), and it explains these traits by appeal to genetic selection. It is thus important to note that the question is whether such traits are adaptations *for religion* in particular, since everyone involved in the debate is an already an adaptationist in the more general sense: all theories in this field explain their target traits as products of selection, rather than as products of other evolutionary causes. Further, CSR theorists are keen to point out that the universal adaptations they identify not only explain the universality of religious belief, but also the variation observed across cultures. Variation is said to occur because the same general cognitive systems produce different specific beliefs in response to different specific inputs.

 One widely cited theory of this kind holds that people believe in supernatural agents because their capacities for detecting agents in their environment are hypersensitive. Psychologist Justin Barrett (2004) has dubbed this trait the *hypersensitive agency detection device*, or HADD, and it is the hypersensitivity itself that is said to be a genetic adaptation. This is because it is assumed that selection would favor agency detection systems that are biased toward false positives over false negatives; it’s better to jump to the false conclusion that there is an agent nearby than it is for there to actually be an agent nearby that one hasn’t detected. As a result, it is claimed, we are genetically predisposed to believe in the existence of agents on the basis of scant evidence. Thus, the universality of supernatural agent beliefs is explained by the fact that individuals across cultures share the same history of genetic selection for the HADD. That the particular beliefs formed vary across cultures is explained by the fact that different environmental contexts trigger different specific beliefs about supernatural agents.

 Of course, the HADD itself cannot explain everything about the formation of religious beliefs. For one thing, religious beliefs are not simply the idiosyncratic beliefs of particular individuals, but are shared widely across individuals. Hypersensitive agency detection does not explain this, but another influential CSR theory does. Anthropologist Pascal Boyer (2001, 2003) explains why some supernatural beliefs spread through social groups better than others do, by appealing to innate and universal patterns of memory and inference. Boyer notes that religious representations typically contradict our ordinary intuitions, and he argues that this makes them stand out as particularly interesting and memorable. Representations of a talking bush, for example, are more salient and compelling than representations of an ordinary bush, and this, Boyer suggests, explains why individuals are more prone to think about, talk about and remember representations of talking bushes. However, if counterintuitive ideas are *too* counterintuitive, then they will simply be incoherent, difficult to process, and difficult to remember. Hence, the spread of successful religious ideas is explained by their being “minimally counterintuitive” (MCI). On this account, what natural selection has actually favored has not been the formation of *religious* beliefs per se, but rather the formation of beliefs about agents in general. So the supernatural agent concepts that are characteristic of religion are said to be formed as by-products of cognitive processes that are adapted to perform other, non-religious functions.

 The HADD theory and the MCI theory explain different cognitive traits using the same general form of explanation, according to which what is adaptive about the traits in question has nothing to do with what is religious about them. This general form of explanation is what I refer to as the “by-product theory,” or the “by-product approach.” But these two traits are not the only features of the religious phenotype to which this approach has been applied. Lawson and McCauley (1990) explain the cognition involved in religious rituals by appealing to universal mechanisms for representing intentional action in general, so, as in HADD theory, beliefs about supernatural agents are just by-products of cognitive systems the function of which is to negotiate interactions with ordinary, natural agents. Bloom (2007) suggests another such trait, observing that religious beliefs all over the world are based on folk intuitions about mind-body dualism, and arguing that these beliefs are “a natural by-product of the fact that we have two distinct cognitive systems, one for dealing with material objects, the other for social entities.” Again, while this view appeals to selection for cognitive systems that happen to produce certain religious beliefs, it does not posit selection for religious belief itself.

 By-product theorists do not assume that the by-product approach offers a *complete* account of religious belief all by itself. For example, while McCauley enthusiastically endorses both the HADD and MCI theories, he also notes that the joint activity of these traits would also explain beliefs about Santa Claus, leprechauns and fairies. “So,” he concludes, “this is only part of the story, but it is a very important part” (McCauley 2004). Bloom notes, more generally, “It is obvious that *some* religious beliefs are entirely learned—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.” Accordingly, it is more fair to characterize proponents of the by-product approach as holding that explanations of this general form explain *many* religious traits, including many that vary across cultures.

**3. Adaptationist Theories of Religion**

3.1 *The General Form of the Adaptationist Approach*

On the other side of the debate, we do not find adaptationists about religion denying that religious beliefs are minimally counterintuitive, or denying the existence of a HADD. Indeed, adaptationists have had little to say about the cognitivesystems and processes that serve as proximate causes for religious belief-formation. They have focused, instead, on traits of social motivation. And while there are several different adaptationist theories on offer, which explain several different specific traits, there is striking and consistent agreement among them all regarding the general source of adaptive value for those traits. Adaptationist theories all agree that what is adaptive about religion is the role it plays in motivating cooperation, especially when cooperation is costly.

 As adaptationists have noted (Atran 2002; Henrich and Gervais 2010; Norenzayan 2013, p. 161), what a purely cognitivist perspective cannot explain is the role of *commitment* in religious belief. Anyone may merely entertain, hypothetically, the content of anyone else’s religious system, and this shows that one may engage in the cognition of a given religion without actually believingit. For a person to be religious, the relevant forms of cognition must carry motivational force. Thus, in contrast to the cognitivist focus of by-product theories, the role of motivation is central to adaptationist theories. These theories recognize that motivations to cooperate often come into conflict with motivations to *not* cooperate, and to pursue self-interest instead. Accordingly, they place special emphasis on the adaptive value of *prosocial* or *altruistic[[2]](#footnote-2)* cooperation. For example, Norenzayan and Shariff (2008) take as their target explanandum not religion in general, but “religious prosociality, or the idea that religions facilitate acts that benefit others at a personal cost.” They begin by noting that, “texts of all major religions explicitly encourage prosociality in their adherents.”

 Obviously, not all religious systems posit an omnipotent God who is concerned with human moral behavior, but the God of Christianity, Islam and Judaism has more followers than any other, and these religious systems conform beautifully to the general form of the adaptationist account: selection for cooperation leads to selection for beliefs and practices that promote prosociality. The same reasoning also applies, however, to other prominent forms of religion. Cooperation could also be increased, for example, by beliefs about ancestor spirits who are far from omniscient, but who still know when believers have violated social norms, and who then dole out rewards and punishments in the form of “fortune” and “misfortune.” The Hindu and Buddhist principle of karma even posits supernatural rewards and punishments that aren’t doled out by a supernatural *agent* at all—a supernatural *force* is sufficient, since beliefs about that force function to motivate compassion for others.

 Thus, while by-product theories take as their target explananda the proximate *causes* of religious belief, adaptationist theories focus, instead, on the motivational *consequences* of such beliefs. It is one thing to entertain a myth from a foreign culture about an all-powerful being who knows what people do and think, and who doles out punishments and rewards accordingly. It is quite another thing to actually believe that such a being is monitoring one’s own thoughts and actions, and to act against one’s own interests as a result. And it is only by engendering such motivations, adaptationist theories claim, that supernatural beliefs become adaptive.

3.2 *Where Adaptationist Theories Differ*

Like the by-product theory, then, the adaptationist theory is in fact a general form of explanation shared by a family of more specific theories, each of which explains different specific traits of the religious phenotype. Unlike by-product theories, however, when adaptationist theories go to specify what the relevant traits are for*,* they do not all appeal to the same kind of selection process. And while adaptationist theories all focus on cooperative behavior, different theories focus on different traits under this heading, depending on the kinds of selection to which they appeal.

 In addition to defending the adaptationist approach to religion in general, Sosis has also done important empirical work supporting the more specific theory of *costly signaling* in religious rituals (Alcorta and Sosis 2005; Sosis and Bressler 2003). This theory, originally put forward by Irons (2001), begins by noting that collective enterprises generate a temptation for individuals to “free ride” on the efforts of others. The more everyone else invests in a collective project, the greater the temptation becomes for each individual to selfishly leave the costs of cooperating to others. “If three hundred other guys are already going out to fight,” the free rider reasons, “what difference will it make if I just sit this one out?” This temptation poses a problem for adaptationist theories of cooperation, because if cooperative motivations can be exploited in this way, then cooperation may often reduce fitness, rather than increasing it. The costly signaling theory addresses this problem by proposing that religious rituals enable genuine cooperators to exclude free riders from their cooperative interactions.

 A dramatic example of costly signaling comes from Hindu devotees in Malaysia, who carry “burdens” in an annual procession to celebrate a mythical victory of good over evil. These burdens often involve piercing the body with thick metal pins, and some participants even suspend their whole bodies in mid-air by hooks pierced through the skin of their backs and legs. Anyone willing to do *that*, it seems, is capable of putting their religious commitments above self interest. If those commitments are widely known to include normative principles that discourage free riding and promote altruism, then publicly displaying burdens in this conspicuous manner can function as a way of credibly signaling to others that one is not a free rider. This makes costly signalers more attractive cooperative partners than non-signalers, providing them with greater benefits from cooperation. As a result, the benefits of costly signaling can outweigh its costs, and these costly behaviors can increase fitness in spite of their costs.

 This analysis could be used to explain how religion evolved by genetic selection at the individual level, but it is also consistent with other forms of selection. Sosis points out that selection for costly signaling may occur at the group level, rather than at the individual level (Sosis and Alcorta 2003), since free riding reduces fitness for groups as well as for individuals. The greater the proportion of free riders in an army, for example, the smaller the number of soldiers who will actually show up to pay the costs of fighting, and the poorer the whole group’s chances of success in military conflict. The same logic also applies, however, for less direct forms of group competition. Many forms of hunting and farming require a great deal of cooperation, and groups full of free riders will generate less food than groups full of genuine cooperators.

 Since free riding is a special form of selfishness, the group-selectionist account of costly signaling is just a special case of the more general principle on which all group-selectionist accounts of religion are founded. These theories argue that while the costs of altruism accrue to individuals, they may be outweighed by benefits that accrue to the social group as a whole. As a result, the same altruistic behavior that reduces the fitness of the altruist relative to other individuals may nevertheless increase the fitness of the altruist’s group relative to other groups. Consequently, altruism can be selected for in spite of its costs. And if religion promotes altruism, then it, too, may be selected for at the group level.

 This account of religious evolution was first developed by biologist David Sloan Wilson (2002), but it has recently been developed further by psychologist Jonathan Haidt (2012), who integrates it into his own theory of moral psychology. And in addition to appealing to group selection, Wilson and Haidt also appeal to cultural selection, or selection among culturally inherited traits (cf. Richerson and Boyd 2005). One reason for this is already hinted at in Bloom’s admission, above, that genetic inheritance does not explain how individuals acquire the particular beliefs and practices of their particular faith. Even when children inherit the religious beliefs of their own parents, they inherit them culturally, rather than genetically; children learn the beliefs and practices of their religion in the process of acculturation, just as they learn to say “please” and “thank you.” The fundamental insight of the theory of cultural selection is that this kind of learning may be treated as a form of non-genetic inheritance, making it possible explain cultural evolution in terms of Darwinian selection.

 To appeal to both cultural evolution and genetic evolution in the same group-selectionist theory, however, is to introduce a critical question: is group selection supposed to explain genetic evolution, or cultural evolution? That is, how are the group-selected traits inherited—by learning, or by genes? To recognize two different forms of inheritance is to recognize two very different types of selection process, and thus two very different forms of adaptationist explanation. Moreover, cultural evolution and genetic evolution interact causally, in what Richerson and Boyd (2005) have dubbed *gene-culture co-evolution*, making it all the more important to identify exactly which type of selection is occurring at which level. Wilson and Haidt claim that, because of gene-culture co-evolution, both genetic group selection and cultural group selection play some role. But this brings their theory into conflict with an alternative adaptationist theory based strictly on cultural group selection (Norenzayan 2013; Atran and Henrich 2010). On this view, group selection acts only on culturally learned traits, not on genetic traits.

 Thus, while we find by-product theories appealing only to genetic selection at the individual level, we find adaptationist theories appealing, in addition, to various forms of cultural evolution and group selection. As I argue in the next section, this difference leads to very different views regarding what counts as an adaptation, and this is one important source of the misunderstanding that has been mistaken for an empirical debate.

**4. Two Different Meanings of ‘Adaptation’**

Since traits inherited in either of two ways (genetic vs. cultural) may be selected for at either of two levels (group vs. individual), the theoretical framework employed by adaptationists about religion recognizes no less than four distinct types of selection process that may be invoked to explain a given trait. By contrast, the framework on which by-product theories are based recognizes only one of these four types: genetic selection at the individual level. I thus distinguish between *broad* and *narrow* forms of adaptationist framework, and this distinction results in an important difference between the definitions of “adaptation” at work in adaptationist theories and by-product theories, respectively.

 Since broad adaptationists recognize four distinct forms of selection process, there is a big difference in these accounts between a trait’s being a product of *selection in general* and it’s being a product of *individual-level genetic selection*, in particular. On this view, a trait counts as an adaptation as long as it is the product of *any* form of selection process. So, for example, for a broad adaptationist it would not follow from the fact that a trait is an adaptation that it is inherited genetically. For narrow adaptationists, by contrast, *all* selection is genetic selection, so to abstract away from facts about genetic inheritance would be to abstract away from facts that are absolutely essential for determining whether a trait can be an adaptation or not.

 If there is a genuine debate underlying the apparent debate between adaptationist and by-product theories of religion, I suspect it lies here, between those who are willing to recognize non-genetic traits as products of selection, and those who insist that selectionist explanations are legitimate only for genetically inherited traits. But if this is indeed a genuine point of conflict between broad and narrow adaptationists, it is important to see that it is not a conflict about religion, specifically, and it is not an empirical debate about how any particular trait or set of traits evolved. Rather, it is a basic methodological debate, concerning the role evolutionary theory plays in explaining human cognition and behavior in general.[[3]](#footnote-3)

 In one sense, then, the broad adaptationist concept of adaptation is a dramatic departure from the ordinary use of ‘adaptation’ in evolutionary theorizing. The concept of adaptation is widely employed today because of its fantastic success in biology in general, and that success is overwhelmingly due to theories based on individual-level genetic selection. Thus, from the point of view of traditional biology, the idea of a “cultural adaptation” or a “group-level adaptation” might sound outlandish. Yet while all human activities are biological phenomena, some of our activities—including religious ones—are biological phenomena that don’t occur in any other species. As a result, adaptationists about religion are attempting to apply biological concepts in a new and special domain, and this has led them to stray from traditional terminology. While such terminological shifts do risk causing some miscommunication, there is nothing unreasonable about adopting a broader definition of ‘adaptation.’ This is because it is reasonable to assume that a trait counts as an adaptation as long as there is something it has been adapted *for*—as long as it possesses an evolved function. Narrow adaptationists already agree with broad adaptationists that traits’ functions are determined by their histories of selection, so broad adaptationists are simply applying this shared assumption to the new, additional forms of selection to which they appeal: any trait that is a product of any kind of selection can be said to have a function. Whether it is a trait of groups or of individuals, and whether it is inherited culturally or genetically, a trait counts as an adaptation as long as *some* form of selection explains its functional properties.

 Adaptationists about religion are sometimes quite explicit about this. David Sloan Wilson (2002) borrows the term “Darwin machine” from Plotkin (1994), and uses it to refer to the general logic of selection processes. He then advocates an ultra-liberal application of adaptationist reasoning, arguing that not only cultural selection but even ordinary rational decision-making (in which an agent selects among options) are examples of Darwin machines. He complains that “It is unfortunate that evolution is so often associated with genetic evolution, a slow process that gives the impression of an incapacity for change over the timescales that matter most to living people struggling with their problems. When we expand our view of evolution to include all Darwinian processes, we can begin to see how religions actually produce transformative change, even from a purely evolutionary perspective” (p. 35).

 Philosophers Russell Powell and Steve Clarke, who have recently argued against the by-product theory, also insist that “the process of adaptation need not be tied to genetic transmission” (Powell and Clarke 2012). Indeed, they offer a concise statement of the central assumptions of what I am calling broad adaptationism, appealing directly to the “cultural group” as a fitness-bearing unit of selection:

There are many clear-cut examples of non-genetic adaptations that increased the fitness of cultural groups, including the manufacture of fire (Wrangham 2009) and the construction of sea-faring vessels (Richerson and Boyd 2005), as well as the development of moral norms, military hierarchies, and sophisticated social exchange networks that were built gradually by ‘invisible hand’ mechanisms that are closely analogous to paradigmatic natural selection (Sterelny 2007). Religion may be a less functionally obvious case than these, but there is nothing inherently problematic about it being an adaptation of cultural groups.

Like Wilson, Powell and Clarke explicitly insist that it makes no difference whether the particular selection process involved is cultural evolution or genetic evolution, or whether it occurs at the level of groups or of individuals. Any product of any selection process is an adaptation.

 These are conceptual claims about what counts as an adaptation, not empirical claims about how any particular traits actually evolved. And while I agree that there is nothing “inherently problematic” about treating the traits of cultural groups as adaptations, it is nevertheless clear that this is not how the concept of adaptation is used in by-product theories. Because adaptationist theories of religion are all based on the broad adaptationist concept of adaptation, when they assert that religion isan adaptation, what they are actually committed to is the assertion that religion is a product of *some kind of selection or other*. By contrast, because they employ a the more narrow, traditional definition of adaptation, when by-product theories assert that religion is *not* an adaptation, what they are actually committed to is the claim that religion is not a product of *individual-level genetic selection, in particular*. But this means that the by-product theory does not deny what the adaptationist theory asserts, because to deny that a trait is a product of one *particular* type of selection is obviously not to deny that it is a product of *any* type of selection whatsoever. To deny that a trait is a product of genetic selection at the individual level still leaves open the possibility that it is a product of cultural selection, group selection, or both.

 Once the empirical commitments of these theories are separated from their more general methodological assumptions, it becomes clear that the by-product theory and the adaptationist theory are mutually consistent. What they disagree about is not how any of the traits of the religious phenotype actually evolved, but rather whether traits that evolved by group selection or cultural selection are deserving of the label ‘adaptation.’

**5. Two Different Meanings of ‘Religion’**

These differing concepts of adaptation provide one way of showing that adaptationist and by-product theories do not conflict, but there is also another way. Just as ‘adaptation’ has different definitions in these two types of theories, so too does ‘religion.’ When by-product theories explain religion, what they actually explain is something much more specific: cognitive traits that are inherited genetically, that are selected for at the individual level, and that are responsible for forming representations of supernatural agents. Likewise, when adaptationist theories explain how religion evolved, they too explain something much more specific: motivationaltraits that may be inherited either by genes or by learning, and that may be selected for at either the group level or the individual level. But since these theories explain these different types of traits, they don’t offer competing explanations for the same empirical facts.

 This is not to suggest that adaptationist theories have nothing at all to say about cognitive traits, of course, since the contents of religious beliefs in these accounts remain essential to explaining cooperative behavior. The point is, rather, that cognitive traits are secondary in an important sense: they are posited only in virtue of their contributions to prosocial motivation. Beliefs about God’s moral attitudes may be one way to trigger the relevant motivations, but entirely different forms of cognition, such as non-theistic beliefs about karma, may do the job as well. As long as a given cognitive trait produces the right motivational responses, adaptationist theories have everything they need. As a result, while by-product theories and adaptationist theories both offer explanations for various cognitive traits, they do not offer explanations for the same kinds of cognitive traits. By-product theories address the proximate *causes* of religious cognition, while adaptationist theories address the motivational *consequences* of this cognition. This opens up the possibility that the very same belief might have both the causes attributed to it by the by-product theory and the consequences attributed to it by the adaptationist theory.

 Suppose, for example, that the HADD causes someone to believe the stories she’s heard about an invisible ancestor spirit who roams the village, punishing bad behavior when she sees it. Then suppose the person has an opportunity to steal something, and chooses not to in order to avoid this supernatural punishment. In such a case, the would-be thief pays an opportunity cost, passing up the benefits of stealing. At the same time, the would-be victim benefits from this decision not to steal, making this exactly the sort of altruistic behavior that, according to adaptationists, makes religion adaptive. Thus, even if the HADD did not *originally* evolve to perform any specifically religious function, once it had *already* evolved it might subsequently happen to have the motivational consequences to which adaptationist theories appeal, and it could then be selected for in virtue of those consequence. In other words, a trait that originally evolved to perform a non-religious function might subsequently acquire a religious function. In such a case, the respective definitions of ‘religion’ in by-product theories and adaptationist theories would not pick out the same facts about the evolution of the same trait. Relative to what the by-product theory actually explains when it is used to explain religion, ‘religion’ in this case would refer to a trait that possesses a non-religious function: the original function of the HADD. By contrast, given what adaptationist theories mean by ‘religion,’ the term would refer here to the specifically religious function that the HADD acquired later on.

 Here we see how systematic differences in the methodological commitments of adaptationist and by-product theories produce systematic differences in the types of facts that each purports to explain. Because the religious phenotype is a complex natural phenomenon, researchers must first analyze it into more manageable constituent traits, such as the HADD, or costly signaling behaviors. But instead of leading adaptationists and by-product theorists to offer competing explanations for the same traits of the religious phenotype, these methodological differences lead them to focus on different sorts of traits. So just as differences in the meaning of ‘adaptation’ render the theories mutually consistent, so too do differences in the meaning of ‘religion’—the target explanandum. Because the theories do not explain the same target facts, they do not provide competing, conflicting explanations.

 This concludes my argument for the consistency of these theories, but there is more to be said about their relationship than merely that they are consistent. If they explain different constituent features of the same complex phenomenon, then how, exactly, are these approaches related? The Goldberg Exaptation Model answers this question.

**6. The Goldberg Exaptation Model**

The Goldberg Exaptation Model (GEM) is a conception of the religious phenotype that I have extracted from the existing literature, because it has been proposed independently by theorists from both sides of the adaptation/by-product debate. It is not a synthesis or hybridization of the two theories, but rather an account of the intersection that already exists between them, a set of shared assumptions about the explananda of evolutionary theories of religion. These assumptions, I argue, are already sufficient to show how adaptationist and by-product theories may be integrated.

 While McCauley was first to propose the GEM, he did so in metaphorical terms that can be more easily understood against the background of a literal description. I thus begin with Sosis’s version. Both theorists present the model in the course of defending their own side of the debate, so their rhetorical aims are not just different, but opposing. Nevertheless, the substance of their descriptive claims reveals that there is nothing more at stake than a matter of emphasis.

*6.1 Sosis’ Mereological Analysis*

Sosis presents the central features of the GEM while accusing by-product theories of committing a sort of mereological error—explaining the parts when they should be explaining the whole. I disagree that this analysis actually shows the by-product theory to be in any kind of error, because I disagree that by-product theories ought to be explaining the whole in the first place. But I agree that the by-product approach only explains the parts, not the whole, and this is the crucial point.

 Sosis notes that the religious phenotype is a complex phenomenon, an “adaptive complex” analyzable into many distinct constituent traits. He also observes, correctly, that what by-product theories explain are the functions of various constituent traits, rather than the function of the complex as a whole. He refers to the whole complex as “the religious system,” arguing that “It is the religious system, not the constituent parts, that produces functional effects and is the appropriate unit of an adaptationist analysis. A proper byproduct account of religion, which has yet to be offered, must explain why the religious system’s constituent parts recurrently coalesce across cultures” (2009).

 In evaluating this criticism, we must distinguish between two different claims. One is the claim that the by-product theory does not explain the whole, only the parts. A different claim, however—implied by the terms “proper” and “appropriate”—is that the by-product theory *ought* to explain the whole. I fully agree with the first claim, and this is ultimately the more important point. But I disagree with the second claim. Granting that by-product theories do not focus on explaining the system as a whole, what would be improper about this? Why think that the appropriate unit of analysis for the by-product theory is the whole, rather than the parts? Certainly *some* theory ought to explain the whole system, but why should this obligation fall to the by-product theory, in particular?

 All theories have limits; they explain what they explain, but they don’t explain everything. Accordingly, it is no fault, flaw or error of a theory that it does not explain facts that fall outside its scope. No form of the theory of natural selection explains why thunder follows lightning, but this does not show that there is anything wrong with this theory. The same can be said here of the by-product theory. Facts about the religious system as a whole simply *fall outside the scope* of the by-product theory. What is important about Sosis’s mereological analysis is that it clearly identifies important facts that only the adaptationist approach can explain, thus illustrating an important sense in which the religious phenotype is an adaptation. But none of this shows that there is anything improper about by-product theories. Rather, what it shows is precisely that the religious system as a whole is *not* the appropriate unit of analysis for the by-product theory; it is the appropriate unit of analysis for *adaptationist* theories, not by-product theories. What emerges from Sosis’s mereological analysis is a natural division of theoretical labor: while by-product theories focus on explaining important parts within the religious system, adaptationist theories focus, instead, on explaining how those parts fit together to form a functionally integrated system—the religious phenotype as a whole.

 This is not to suggest that the target facts for one approach are inherently off limits to the other. Rather, it is just to say that there are two different explanatory jobs to be done, and two different tools to do them with, and while the by-product approach is well suited to one of them, the adaptationist approach is still needed for the other. It may well be that adaptationist theories also explain certain parts, in addition to explaining the whole, if some parts first evolved as parts of the system from the very beginning. But as long as *some* significant parts are by-products, and as long as there is also a larger system composed from these parts, both theories have essential roles to play.

*6.2 McCauley’s Metaphor and the Role of Exaptation*

McCauley (2004)describes the religious phenotype in very different terms, but the metaphor he uses maps perfectly onto Sosis’s mereological analysis:

The mind does not contain a specific department of religion. Instead, religion exploits a diverse collection of emotional and cognitive inclinations in human beings that enjoy neither logical nor psychological unity. The upshot of this analysis is that *cognitively speaking* religion is a Rube Goldberg device, which is to say that it is an exceedingly complicated contraption calling on all sorts of psychological propensities that are, otherwise, usually unlinked. (emphasis in the original)

This quote appears in an article endorsing the by-product approach, and McCauley emphasizes the task of explaining the independent, “otherwise unlinked” parts of the religious system. Nevertheless, he also recognizes that these parts have subsequently been linked together into some larger system, or “contraption.” A Goldberg device has a distinct function of its own, over and above the independent, prior functions of its constituent parts. Indeed, the entertaining charm of Goldberg devices lies precisely in the ironic bad fit between the general function of the whole device and the specific functions of its disparate parts. As a result, although McCauley uses the metaphor to highlight the special contribution of the by-product theory, it could just as easily be used to highlight the special role of the adaptationist theory instead. Armed with this image, Sosis could point out that while by-product theories do explain the prior functions of various parts, what they still don’t explain is the general, overarching function of the contraption itself.

 Likewise, however, while Sosis emphasizes the importance of the whole, he also recognizes the need to explain the parts. He makes this point by appealing to the concept of *exaptation*[[4]](#footnote-4):

The most likely evolutionary scenario is that cognitive, emotional, and behavioral elements were exapted for use in a complex system of communication, cooperation, and coordination, namely the religious system. An exaptation is a preexisting trait that acquires a new role for which it was not originally designed by natural selection (Gould and Vrba 1982). Importantly, exaptations have functional effects but exapted traits are not modified when taking on their new role; if they are, adaptive modifications are known as secondary adaptations.

By explicitly describing constituent traits in the system as exaptations rather than secondary adaptations, Sosis implicitly recognizes that the original, non-religious functions of certain parts are nevertheless important. While still retaining their non-religious functions, exapted traits of the religious phenotype acquire new functional roles *in addition*, and these are the roles they play within the Goldberg device. But in order to explain how non-religious traits ultimately acquired religious roles, it’s necessary to understand what their original, non-religious roles were to begin with.

 Thus, while Sosis and McCauley differ in emphasis, both agree with the following three claims, which comprise what I am calling the Goldberg Exaptation Model:

(1) The religious phenotype is a complex phenomenon composed of more specific constituent traits.

(2) At least some constituent traits originally evolved to perform non-religious functions.

(3) Afterthey had evolved, these traits were exapted by selection to play new, additional roles within the larger system of the religious phenotype—a system that possesses its own distinct function.

While (2) gives by-product theorists what they insist upon, (3) gives adaptationists what they insist upon. If important traits of the religious phenotype originally evolved to perform non-religious functions, then those traits count as by-products in the sense that by-product theorists insist upon. But this does not entail that they are *only* by-products. For if selection *also* explains why, after they initially evolved, the same traits later acquired additional functional roles as parts of a religious system, then the system itself counts as an adaptation in the sense that adaptationists insist upon. And if it is the system as a whole, rather than any of its parts, that is responsible for the common-sense distinction between religious individuals and non-religious individuals, then adaptationists seem to be within their rights in declaring the system as a whole an adaptation *for religion*.

**7. Unifying By-product and Adaptationist Accounts**

Having thus explained what I take the GEM to be, I now want to illustrate, by example, how it can be used to integrate by-product accounts with adaptationist accounts. I approach this by considering a potential counterexample to my claim that adaptationist and by-product theories explain different facts. We’ve seen already that Bloom (2007) offers a by-product account of the trait of folk dualism, according to which representations of minds without bodies result from the “incommensurable outputs” of two distinct cognitive systems: one that evolved for representing the movements of ordinary physical objects, and another that evolved specifically for representing the behavior of agents with minds. However, psychologist Jesse Bering (2006) also offers an adaptationist account of folk dualism, making this one trait for which both a by-product account and an adaptationist account have been offered. As a result, these two accounts of folk dualism come the closest to manifesting a genuine empirical disagreement.

 Considering this case in light of the GEM, however, reveals that they do not, in fact, provide competing explanations. For while they do explain facts about the same trait, they do not explain the same facts about that trait. Indeed, the facts they explain may well be separated by millions of years of evolutionary history. The GEM thus shows how both theories make distinct contributions to the same causal narrative.

 On Bering’s account, selection for folk dualism is due to the role this trait plays in helping people manage their reputations. If the bodiless minds being represented happen to belong to supernatural agents who punish antisocial behavior and reward prosocial behavior, then, Bering suggests, dualistic cognition will cause people to behave in ways that contribute positively to their reputations. This is assumed to increase fitness, because humans’ social relationships are crucial for their fitness, and good reputations are crucial for maintaining good social relationships. A bad reputation can make it hard to attract cooperative partners, and a *very* bad reputation can be downright dangerous, even deadly. Thus, like other adaptationists, Bering assumes that the general function of the religious phenotype is to motivate costly forms of cooperative behavior, and what makes dualistic cognition an adaptation for religion is the fact that selection for dualistic concepts depends on the impact they have on social motivations. That is, dualistic beliefs are adaptations for religion specifically in virtue of the contribution they make to the general function of the religious phenotype as a whole.

 On Bloom’s account, by contrast, the selection pressures that explain the function of dualistic cognition have nothing to do with motivating cooperation. Rather, the claim is that selection favored those of our ancestors who represented agents using a distinct and specialized system for agents, rather than those who attempted to represent the behavior of agents (e.g., predators) using the same cognitive processes used for representing non-agentive objects (e.g., rocks). The separation between these two systems makes it possible for representations of agents to be triggered in the absence of any representations of those agents’ bodies, and these universal cognitive traits cause dualistic concepts to show up frequently across different religious systems. But this is not because selection favored specifically religious forms of dualistic belief.

 For the sake of argument I will assume that both accounts of our evolutionary history are accurate. What is important to note is that the ecological conditions to which Bloom appeals would already have been in place a *very* long time ago—even by evolutionary standards. The selection pressures that explain why agent-specific cognition is adaptive are not recent changes in the ecology of our lineage. Indeed, they would have been in place long before our ancestors resembled anything like a human being, and probably long before they resembled mammals. Our distant reptilian ancestors would have had every bit as much reason as our recent primate ancestors did to represent predators and mates using special cognitive systems that treat agents differently from non-agents. Clearly, a trait that first evolved among our reptilian ancestors would not be a trait that evolved to perform a specifically religious function. But there is room in evolutionary history for a lot to happen afterthese two distinct systems first evolved.

 What Bloom’s account does not explain, as Sosis points out, is why beliefs about bodiless agents should recurrently exhibit certain functional interactions with other psychological traits, such as motivations to pay the costs of following social norms. Bloom’s account does not explain why, across different religious systems, we recurrently observe dualistic beliefs of a very specific kind: those in which bodiless agents care about human social life, possess supernatural powers, and use their powers to punish selfish behavior and reward altruistic behavior. Back when members of our lineage were still reptiles or four-legged mammals, dualistic beliefs with these particular properties would have held no adaptive value. But much later on, as our ancestors began to resemble modern humans, selection pressures would have changed in relevant ways. For example, only after language evolved would individuals have had the ability to share third-party information with each other about the behavior of others in the community. The evolution of language thus created new selection pressures for the effective management of reputation. And as Bering argues, these new pressures would explain specifically religious forms of dualistic thinking, and specifically religious forms of psychological interaction between folk dualism and other traits. Bering’s account thus explains what Bloom’s account does not: how the trait of folk dualism was exapted to perform the particular functional role it plays within the religious phenotype.

**8. Conclusion**

I hope to have shown two things. The first is that a careful examination of the theories in question reveals them to be mutually consistent. Adaptationist theories and by-product theories don’t disagree about whether religion is an adaptation, because they don’t take the terms ‘religion’ and ‘adaptation’ to refer to the same empirical phenomena. As a result, the fact that they provide opposing answers to the question “Is religion an adaptation?” does not show that they provide conflicting, competing explanations.

 The second thing I hope to have shown is that there already exists in the literature a set of shared assumptions capable of unifying the two theories. The Goldberg Exaptation Model brings to light shared assumptions about the explananda of theories of religious evolution that have already been endorsed independently by theorists from both camps. These assumptions reveal the nature of the basic relationship between the two theories: while the by-product approach is well suited to the task of explaining original, non-religious functions of various constituent traits, it is not capable of explaining how those traits were assembled together to form a coherent “device” possessing a function of its own. That task, however, is one to which the adaptationist approach is perfectly suited.

 It must be noted that certain adaptationists, who emphasize the role of cultural selection in religious evolution, have already recognized the potential for unification. While they have made no effort to diagnose the source of all the debate and disagreement, and while they do not employ the terms of the GEM, the positive account they provide is one in which cultural evolution has cobbled together a Goldberg device using the by-products of genetic evolution. Anthropologists Scott Atran and Joseph Henrich say of their theory: “This synthesis integrates insights from studies of the cognitive foundations of religion with evolutionary approaches to human cooperation to derive a deeper understanding of the origin and development of prosocial religions” (2010). Likewise, psychologist Ara Norenzayan takes integration to be a major aim of his theory: “The argument in this book is an attempt at integrating these two perspectives—the social and the cognitive—that are currently seen as competing accounts” (2013, p. 11). He ultimately concludes that “successful religious groups—the cultural ancestors of most human beings alive today—pieced together, step by step, a whole cluster of psychological mechanisms that, building on supernatural monitoring and credible displays of sincere faith, fostered and cemented social solidarity” (2013, p. 162). I fully endorse these calls for integration and unification, of course, and I offer the Goldberg Exaptation Model as a way of showing how they can be met.

**Acknowledgments**

I would like to thank Richard Sosis, John Beatty, Ara Norenzayan, Christopher Stephens and Eric Margolis for important comments on earlier drafts.

**References**

Alcorta, C. S., and R. Sosis. 2005. “Ritual, Emotion, and Sacred Symbols.” *Human Nature* 16 (4): 323–359.

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

Atran, S., and J. Henrich. 2010. “The Evolution of Religion: How Cognitive by-Products, Adaptive Learning Heuristics, Ritual Displays, and Group Competition Generate Deep Commitments to Prosocial Religions.” *Biological Theory* 5 (1): 18–30.

Barrett, J. L. 2004. *Why Would Anyone Believe in God?* Plymouth: AltaMira Press.

Bellah, R. N. 2011. *Religion in Human Evolution: From the Paleolithic to the Axial Age.* Cambridge, MA: Harvard University Press.

Bering, J. M., and D. D. P. Johnson. 2005. “‘ O Lord You Perceive My Thoughts From Afar’: Recursiveness and the Evolution of Supernatural Agency.” *Journal of Cognition and Culture*, 5 (1): 118–142.

Bering, J. M. 2006. “The Folk Psychology of Souls.” *Behavioral and Brain Sciences* 29 (5): 453–461.

Bloom, P. 2007. “Religion Is Natural.” *Developmental Science* 10 (1): 147–151. doi:10.1111/j.1467-7687.2007.00577.x.

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

Boyer, P. 2003. “Religious Thought and Behavior as by-Products of Brain Function.” *Trends in Cognitive Sciences* 7 (3): 119–124. doi:10.1016/S1364-6613(03)00031-7.

Bulbulia, J. 2009. “Religiosity as Mental Time-Travel: Cognitive Adaptations for Religious Behavior.” In J. Schloss and M. J. Murray (Eds.), *The Believing Primate: Scientific, Theological, and Philosophical Reflections on the Origin of Religion* (pp. 44-75). New York: Oxford University Press.

Durkheim, E. 1995. *The Elementary Forms of Religious Life*. New York: Free Press.

Evans-Pritchard, E. 1956. *Nuer Religion*. Oxford: Oxford University Press.

Evans-Pritchard, E. 1965. *Theories of primitive religion*. Oxford: Oxford University Press.

Gould, S. J., and E. S. Vrba. 1982. “Exaptation: A Missing Term in the Science of Form.” *Paleobiology* 8 (1): 4–15.

Haidt, J. 2012. *The Righteous Mind*. New York: Pantheon Books.

Henrich, J., and W. M. Gervais. 2010. “The Zeus Problem: Why Representational Content Biases Cannot Explain Faith in Gods.” *Journal of Cognition and Culture* 10 (3): 383–389. doi:10.1163/156853710X531249.

Hume, D. 1992. *The Natural History of Religion*. Ed. J. Fieser. New York: Macmillan.

Hume, D. 1993. Dialogues concerning natural religion. In J. A. C. Gaskin (Ed.), *Dialogues and Natural History of Religion.* New York: Oxford University Press.

Irons, W. 2001. “Religion as a Hard-to-Fake Sign of Commitment.” In R. M. Nesse (Ed.), *Evolution and the Capacity for Commitment* (pp. 292–309). New York: Russell Sage Press.

James, W. 2004. *Varieties of Religious Experience.* New York: Touchstone.

Kirkpatrick, L. 2008. “Religion Is Not an Adaptation: Some Fundamental Issues and Arguments.” In J. Bulbulia, R. Sosis, E. Harris, R. Genet, C. Genet, and K. Wyman (Eds.), *The Evolution of Religion: Studies, Theories, and Critiques* (pp. 61-65). Santa Margarita, CA: Collins Foundation Press.

Lawson, E. T. and R. N. McCauley. 1990. *Rethinking Religion: Connecting Cognition and Culture.* Cambridge: Cambridge University Press.

McCauley, R. N. 2004. “Is Religion a Rube Goldberg Device? or Oh, What a Difference a Theory Makes!” *Studies in the History of Religions*: 45–64.

Norenzayan, A. 2013. *The Making of Big Gods: How Religion Transformed Cooperation and Conflict.* Princeton: Princeton University Press.

Norenzayan, A., and A. Shariff. 2008. “The Origin and Evolution of Religious Prosociality.” *Science* 322 (5898): 58–62.

Pinker, S. 2006. “The Evolutionary Psychology of Religion.” In P. McNamara (Ed.), *Where God and Science Meet: How Brain and Evolutionary Studies Alter Our Understanding of Religion* (pp. 1-9). Westport, CT: Praeger Perspectives.

Plotkin, H. 1994. *Darwin Machines and the Nature of Knowledge.* Cambridge, MA: Harvard University Press.

Powell, R., and S. Clarke. 2012. “Religion as an Evolutionary Byproduct: A Critique of the Standard Model.” *The British Journal for the Philosophy of Science* 63 (3): 457-486. doi:10.1093/bjps/axr035.

Pyysiäinen, I., and M. Hauser. 2010. “The Origins of Religion: Evolved Adaptation or By-Product?” *Trends in Cognitive Sciences* 14 (3): 104–109. doi:10.1016/j.tics.2009.12.007.

Richerson, P. J. and R. Boyd. 2005. *Not By Genes Alone*. Chicago: The University of Chicago Press.

Sosis, R. 2009. “The Adaptationist-Byproduct Debate on the Evolution of Religion: Five Misunderstandings of the Adaptationist Program.” *Journal of Cognition and Culture* 9 (3): 315–332. doi:10.1163/156770909X12518536414411.

Sosis, R., and C. Alcorta. 2003. “Signaling, Solidarity, and the Sacred: The Evolution of Religious Behavior.” *Evolutionary Anthropology: Issues, News, and Reviews* 12 (6): 264–274. doi:10.1002/evan.10120.

Sosis, R., and E. Bressler. 2003. “Cooperation and Commune Longevity: A Test of the Costly Signaling Theory of Religion." *Cross-Cultural Research* 37 (2): 211-239.

Sterelny, K. 2007. “SNAFUS: An Evolutionary Perspective.” *Biological Theory* (2): 317–28.

Tylor, E. 1871. *Primitive Culture*. New York: Harper.

Weber, M. 2002. *The Protestant Ethic and the Spirit of Capitalism, and Other Writings.* Penguin.

Wilson, D. S. 2002. *Darwin’s Cathedral: Evolution, Religion and the Nature of Society*. Chicago: University of Chicago Press.

Wrangham, R. 2009. *Catching Fire: How Cooking Made Us Human*. New York: Basic Books.

1. An exception is the account offered by Bering (2006), discussed later on. [↑](#footnote-ref-1)
2. The focus here is on evolutionary altruism, rather than psychological altruism. That is, in adaptationist theories of religion, the costs and benefits of altruism are measured in terms of the *consequences* of behavior for selection, rather than in terms of the mental states that *cause* the behavior. [↑](#footnote-ref-2)
3. As a result, it should be noted that while these questions are more general than those concerning religion, they are nevertheless muchmore specific than another set questions about adaptationism that have been much discussed in the literature. Godfrey-Smith (2001) describes three different forms of adaptationism, and Lewens (2008) expands this taxonomy to include no less than seven distinct forms of adaptationism. The questions addressed by this literature, however, concern the relative importance of adaptationist vs. non-adaptationist explanations in *biology in general*, rather than the relative importance of two distinct forms of adaptationist explanation within the specific domain of *human behavior and cognition*. [↑](#footnote-ref-3)
4. Powell and Clarke (2012) endorse this view as well, and they come closer than Sosis to seeing it as a reason for dissolving the debate, if not as a method for actually integrating the theories. They call their account a “pluralist” view, and they note that some traits may be by-products while others are adaptations. They even note that, “The evolution of religion debate is often cast as a forced choice between byproduct and selectionist theories of religion, but this is a false dichotomy.” Nevertheless, instead of concluding that the debate is empty, they frame their argument as an objection to the by-product approach, and their pluralist view is ultimately indistinguishable from Sosis’s view. They conclude that, “If religion is composed of a suite of co-opted byproducts that were gradually organized and improved over time through cumulative selection, then it looks less like an exaptation and more like an ordinary adaptation.” [↑](#footnote-ref-4)