Are Supernatural Beliefs Commitment Devices For Intergroup Conflict?

Robert Kurzban

John Christner

University of Pennsylvania

Robert Kurzban
Department of Psychology
3720 Walnut St.
Philadelphia PA 19104
(215) 898-4977
kurzban@psych.upenn.edu

Abstract. In a world of potentially fluid alliances in which group size is an important determinant of success in aggressive conflict, groups can be expected to compete for members. By the same token, individuals in such contexts can be expected to compete for membership in large, cohesive groups. In the context of this competition, the ability to signal that one cannot change groups can be a strategic advantage because members of groups would prefer to have loyal allies rather than confederates who might switch groups as conditions change. This idea might help to explain why members of certain kinds of groups, especially competitive ones, use marks, scars and other more or less permanent modifications of their bodies to signal their membership. To the extent that people with these marks have difficulty joining rival groups, these marks are effective in signaling one's commitment. It is possible that the public endorsement of certain kinds of beliefs have the same effect as marks. In particular, there are certain beliefs which, when endorsed, might make membership difficult in all but one group. This idea is proposed as an explanation for supernatural beliefs.

Are Supernatural Beliefs Commitment Devices?

Arguably the most important political event of the albeit still young 21st century was a case of intergroup conflict in which supernatural beliefs played a pivotal role. The attack on the World Trade Center in New York City, the Pentagon in Washington DC, and the foiled attack by the hijackers of United Airlines Flight 93 on September 11th, 2001, was motivated by intergroup conflict, but made possible in no small part because the perpetrators had beliefs about the afterlife. While we do not attempt here to sort out the many causal antecedents of this attack, we do propose an explanation for the broader phenomenon: why people entertain supernatural beliefs and their relationship to intergroup conflict.

Introduction

True beliefs are useful, so much so that philosophers have argued that the only thing that minds are good for is "the fixation of true beliefs" (Fodor, 2000, p. 68), sentiments which have been echoed by others (e.g., Dennett, 1987; Milikan, 1984; for a recent discussion, see McKay & Dennett, in press). The general idea is intuitive and compelling: true beliefs aid in accomplishing goals and, with appropriate inference machines, generating additional true beliefs.

Symmetrically, false beliefs are, in general, less useful. Acting on the basis of beliefs which do not capture something true about the world can lead to any number of bad outcomes. False beliefs about what is edible can lead to poisoning, false beliefs about what is sharp can lead to cuts, and so on. False supernatural beliefs, as Wright (2009) has recently documented, cause their bearers to engage in an array of costly behaviors, including enduring – even self-inflicting – severe harm

and – from an evolutionary standpoint, the most costly choice of all – electing to forgo reproduction. (See also Iannaccone, 1992).

In light of these arguments, one would expect minds – absent some selective force – to be designed to resist adopting false beliefs. There are, however, important exceptions. Consider binary decisions – such as fleeing or not fleeing from a potential predator – in which the costs of errors (misses, false alarms) and the benefits of being correct (hits, correct rejections) are asymmetrical. In such cases, if the system is forced to adopt one belief or other and act on the basis of the belief, selection will *not* favor maximizing the probability of true belief; it will rather maximize expected value (Green & Swets, 1966; Cosmides & Tooby, 1987). That is, if we assume that there must be a belief *either* that the predator is present *or* that it is absent (as opposed to some probabilistic representation), then even weak evidence should give rise to the (likely false) belief that the predator is present so that the appropriate action (i.e., fleeing) can be taken.

This principle is reflected in the design of both human artifacts, such as the smoke detector, and human physiology (Nesse, 2001, 2005; Nesse & Williams, 1994). A smoke detector cannot signal that there might be a fire, so it signals that it "believes" there is one even on scant evidence. In humans, all-or-none defenses such as the immune system (Nesse, 2001) reflect the same idea.

This principle governs the design of evolved mechanisms for inferences about the state of the world across any number of domains. As Wiley (1994) put it, rather than maximize percent correct, "[b]asic decision theory suggests that a criterion should maximize the expected utility for the receiver…" (p. 172). Wiley

shows, using a standard signal detection analysis, that selection can favor "adaptive gullibility" – erring on the side of false positives in the context of mating – and "adaptive fastidiousness" – erring on the side of misses in the context of detecting prey. The propensity for error – false "beliefs" about what is and is not a mate or prey – is built into these mechanisms because selection will sift in design space for designs that maximize fitness rather than accuracy. This is as true for evolved human systems as it is for other organisms' systems. (Haselton & Buss, 2000; Nesse & Williams, 1994; Tooby & Cosmides, 1987).

There is a second important selection pressure that can counteract the tendency for evolution to favor truth-preserving belief systems. This pressure arises in the context of *strategic interactions*, in which individuals' payoffs are affected by others' actions (von Neumann & Morgenstern, 1944; Maynard Smith, 1982). To see the potential advantages of false beliefs, denote p as the true state of the world and p^* as a false belief ($p \neq p^*$). Suppose ego is better off in terms of social advantages if everyone believed p^* rather than p. (Suppose p* is that ego is highly intelligent, for example.) Suppose further that ego, by herself believing p^*, increases the chances that others will adopt p*. (We assume that "genuine" belief can have advantages over simply dissembling, perhaps by virtue of the probability of persuasion; Trivers, 2000) In such a case, *by virtue of the effects p* has on others' behavior, it can be advantageous for ego to believe p** (Nesse & Lloyd, 1992; Trivers, 2000). So-called "positive illusions" (Taylor & Brown, 1988) might be such cases, in which false beliefs about oneself can aid in persuading others to adopt this strategically advantageous belief p* (Kurzban & Aktipis, 2006, 2007). Systems can come to be

designed to generate and adopt *p**'s as long as the costs of the false belief do not outweigh the strategic benefits (Kurzban, in press).

It is important to bear in mind both the power and the limits of this type of argument. Putative cases of design to bring about false beliefs must respect the distinction between, on the one hand, when the decision one makes in and of itself determines one's payoff, and, on the other hand, when the decision one makes *and what one communicates to other agents* affects one's payoff.

The distinction is important because the relentless calculus of decision theory and natural selection punishes mechanisms that do not maximize expected value. *Holding aside what is communicated to another individual* – and thereby potentially changing their behavior and, in consequence, the decision-maker's downstream payoff – a mechanism that maximizes expected value cannot be beaten. (Maximizing expected value is, of course, not the same as maximizing percent correct, as indicated above.)

Substantial confusion surrounds this point. For example, consider the putative benefits of being "too" optimistic. Systems that generate errors that cause one to try more than one "should" – given the expected value of trying or not trying – will lose the evolutionary game to systems that maximize expected value. There is no way around this.

Indeed, recent models that purport to show the advantages of error that do not derive either from the strategic advantages of influencing others' behavior or the decision theoretic advantages captured by signal detection theory succeed only because they artificially penalize strategies that maximize expected value. Nettle

(2004), for example, models a decision in which communication plays no role, so an algorithm that maximizes expected value cannot be beaten by any other strategy without giving non-maximizers help. In the model, "optimists" – who overestimate the chance of success – are given exactly such help: The model's "rational" (non-optimistic) agents rely on and use completely *inaccurate* estimates of the chance of success. When agents have no information at all about the chance of success, they should use the decision-theoretic correct estimate of .5 in making their decision. It is true that when the expected payoff of trying is higher than the expected cost of failing, then "optimists" are better off than the "rational" agents (and symmetrically for pessimists; see also Haselton & Nettle, 2006), but the model's "optimists" and "pessimists" win only because they throw out the misleading information that the "rational" agents do not. As we explore below, one prominent model of supernatural punishment runs into this problem as well.

Outside of cases such as these, as far as we know (McKay & Dennett, in press), in which there is an advantage to error because of considerations of decision theory or the value of the communicative effect of one's decisions, one would not expect to find mechanisms designed to adopt false beliefs. Further, one would expect human computational architecture to be designed to reject false beliefs, given their potential costs.

From this perspective, the fact that humans seem to have mechanisms that endorse supernatural beliefs (SNBs) – which are (by assumption) *guaranteed* to be false – is puzzling. First is the bare fact that humans seem not just disposed, but positively eager to endorse supernatural beliefs (Dawkins, 2006; Dennett, 2006).

Second, these beliefs seem to have high costs. Even holding aside the relationship between supernatural beliefs and intergroup conflict – the subject here – supernatural beliefs seem to play a large role in any number of costly behaviors. This would include things like time-consuming (but useless) prayer, building monuments to non-existent gods, sacrificing goats or other animals without consuming them, doing rain dances, taking risks because of predictions of divine intervention, and so on.

So, holding aside the two arguments above, selection should, *everything else equal*, have eliminated belief-generation mechanisms that had the property of generating and acquiring supernatural beliefs. Why, then, are supernatural beliefs so pervasive in our species?

Theories of Supernatural Belief

Many scholars have addressed the issue of the origin of supernatural belief. Here we discuss only a few prominent models, which, broadly, fall into two classes. The first class is *byproduct* explanations. On this view, humans have mechanisms designed to construct, transmit, and acquire representations for one function, and supernatural beliefs emerge as a side-effect of the way these systems operate. We review these first, and then turn to the second possibility, that the mechanisms that generate supernatural beliefs are designed for precisely this function.

Byproduct views

One of the most prominent byproduct models of SNBs begins with the broad idea that people transmit information socially. People learn from one another in part because there are tremendous cost savings in socially rather than individually

learning information (Boyd & Richerson, 1985). Further, given social learning, it follows that, by virtue of the way that learning mechanisms operate, some kinds of ideas, beliefs, and practices will be more likely to be generated, recalled, and transmitted than others (e.g., Sperber, 1985). This is a natural consequence of any social learning system, and this idea is easily seen in the domain of language, in which various rules constrain the grammar entertained by language learners (Pinker, 1994).

From this, it follows that, by an evolutionary process, certain ideas will tend to persist and be observed over time more than others. Those ideas that are "sticky," having properties that make them memorable and transmitted (Bartlett, 1932), will be observed more than those that do not "fit" with human cognition.

One of the major models surrounding SNBs – the "ontological heresy" (OH) model – begins with this idea, and turns on one important element of learning systems, that there seem to be a set of categories of entities that the mind is prepared to learn about. Each of these categories comes with a set of defining characteristics that apply to all entries within it, so when a new entry is added, many of its features are "automatically" assigned, eliminating the need to relearn them. For example, categories like PERSON, ANIMAL, TOOL, PLANT, or OBJECT each provide a scaffolding of inferences on which to build new concepts. When learning about a new animal, people do not need to relearn that the animal's innards resemble those of conspecifics, that it has offspring which grow into adults, that it moves of its own accord and pursues goals, and so on. These inferences are automatically provided by the ANIMAL category.

The OH model highlights that *SNBs tend to be representations that conform to ontological templates but, crucially, depart from them in a particular way* and that this combination – conformity plus exceptions –gives rise to their "stickiness."

Consider a ghost, which is a PERSON, but violates the usual template in that it passes through objects and, most importantly, is not alive, a critical feature of a PERSON. A ghost, then, can be understood as a PERSON – preserving most PERSON-related properties (has a mind, moves around, etc.) plus violations – a ghost can pass through solid matter while people cannot.

Boyer and Bergstrom (2008) recently wrote about such ideas such as ghosts:

> *Such notions are salient and inferentially productive because they combine specific features that violate some default expectations for the domain with nonviolated expectations held by default as true of the entire domain (Boyer 1994). These combinations of explicit violation and tacit inference are culturally widespread and constitute a memory optimum (Barrett & Nyhof, 2001, Boyer & Ramble, 2001). This may be because explicit violations of expectations are attention-grabbing, whereas preserved nonviolated expectations allow one to reason about the postulated agents or objects (Boyer 1994). (p. 119)*

The key point is the notion of a "memory optimum." On this view, supernatural beliefs persist as a byproduct of the fact that human computational systems "like" representations that allow one to reason about them (the PERSON part of a ghost or spirit) combined with the fact that we attend to ideas that violate

our expectations (the non-living component of being a ghost). SNBs, on this view, persist as a byproduct of mechanisms designed for inferences and attention.

A related byproduct view is that some beliefs, by virtue of their content and their tendency to move from one head to another, replicate themselves not because the beliefs are useful to the people who have them, but simply because they are the sorts of beliefs that lead to their own propagation. Dennett (2006) argues that religious systems of belief seem to have properties that make them good at replicating themselves, including the injunction to transmit information to children, reproduce, and conquer and convert others. These features of a belief system, he argues, contribute to the spread of the beliefs themselves.

There are three primary difficulties of these models. First, as the costs of SNBs increase, so does the strength of selection to "clean up" the system, making byproduct claims less plausible. That is, byproduct explanations are unlikely to the extent that costs are high and selection could have selected out these supernatural-belief-generation systems *without compromising the system that these belief-generation systems are a byproduct of.* We believe that these costs are, indeed, high, and that there is no reason to think selection could not have modified learning systems to resist, rather than endorse, supernatural beliefs. Second, byproduct hypotheses explain why SNBs are memorable, but not why SNBs are endorsed (Dennett, 2006). These are two importantly distinct claims. Finally, models such as Dennett's (2006) rest on largely domain-general and content-free learning systems, which, from an evolutionary view, are unlikely to characterize human psychology (Tooby & Cosmides, 1992).

Adaptationist views

The second class of arguments suggest that the mechanisms that underlie SNB acquisition are *designed* to adopt them. On this view, there is some advantage to having SNBs, and this advantage explains the existence of the mechanisms designed to generate and adopt them.

On prominent account is that SNBs "steered individuals away from costly social transgressions resulting from unrestrained, evolutionarily ancestral, selfish interest (acts which would rapidly become known to others, and thereby incur an increased probability and severity of punishment by group members)" (Johnson & Bering, 2006, p. 219). That is, those with SNBs – particularly false beliefs about punishment – would have avoided actions that would have led to costs in the real world, thus making them better off.

This argument is a game theoretical argument that agents with these SNBs could invade a population of agents without them. In evaluating this argument, the key is to consider a population at equilibrium. This would be a population of agents who maximize expected value. In a world in which some acts are punished, *maximizing expected value entails taking into account the probability of detection and the costs of punishment*. Maximizing individuals do not take advantage of *all* opportunities for selfish, norm-violating gain; they take advantage of opportunities with positive expected value. Johnson and Bering assume this issue away: "As long as the net costs of selfish actions from real-world punishment by group members exceeded the net costs of lost opportunities from self-imposed norm abiding, then god-fearing individuals would outcompete non-believers" (Johnson & Bering, 2006,

p. 219). However, there is no reason to think that the default state is a design that favors engaging in (selfish) actions with negative expected value. Indeed, the reverse is the case. Selection should continuously push computational mechanisms toward such optima, subject to all the usual constraints (see, e.g., Dawkins, 1982). Absent an argument about a constraint that is pushing the design off this optimum, game theoretic models must assume expected value maximization as the default.

Further, even if one were to assume that at some point a population were out of equilibrium in this way, such a population is always invadable – again, by agents who do not adopt outcome-reducing SNBs. If the social world were like poker, consider the cost of having the view that those who bluff will endure endless punishment in the afterlife (and, therefore, never bluff). Such people are at a disadvantage and will lose, eventually, to those who use bluffing as a tactic, unhindered by false beliefs about the costs.

A second adaptationist argument for SNBs turns on the value of such beliefs in the context of signaling to others. Arguments of this nature draw on the behavioral ecology literature, especially models that show that some signals evolve because of, rather than in spite of, their cost (Grafen, 1990; Zahavi, 1975). The typical example is the peacock's tail. Because the large tail has great energetic costs and makes one vulnerable to predation, only very healthy and high quality organisms can support them. For this reason, peahens who select peacocks with such tails as mates are at an advantage.

In the context of religion, it has been argued that enduring the high costs imposed by religions – physical harm, deprivation of food and water, labor

requirements, etc. – send signals to others (Irons, 2001; Sosis & Alcorta, 2003). In particular, has been argued that these costs commit those who endure the costs to the group.

However, care must be exercised in the relationship between cost and signal. In the case of the peacock's tail, the cost conveys something about quality as an intrinsic feature of the cost. Poor quality peacocks simply cannot endure the cost. The same argument does not apply to costs and commitment. *Enduring a cost to enter a group does not, as a consequence of the cost, prevent someone from defecting or leaving the group*. All costs in this sense are sunk, as are costs that are imposed while one is in the group (such as a tithe).

Performing rituals can indeed be costly, and such rituals often include SNBs as justification. Enduring such costs might be signaling something. However, it is not clear that these costs signal commitment, given that it is possible to endure costs and then leave the group. Having said that, some kinds of signals might, in fact, make leaving more difficult. We now turn to this issue, and our own view of the function of supernatural beliefs.

SNBs as Commitment Devices.

The Value of Commitment

Difficulties with existing explanations for SNBs suggests that it might be worthwhile to look for alternatives. The idea sketched here requires several inferential steps, and is therefore perhaps not the most elegant model, but arguably solves the problems with previous models.

We begin with the premise that human evolutionary history was characterized by shifting coalitions and alliances (DeScioli & Kurzban, 2009; Kurzban & Neuberg, 2005; Kurzban, Tooby & Cosmides, 2001; Tooby, Cosmides, & Kurzban, 2003; Sidanius & Kurzban, 2003; Tiger, 1969). This is not to say that some alliances weren't relatively stable, such as those arranged along kin lines, as observed in other species, such as baboons (Cheney & Seyfarth, 2007). The argument turns only on the notion that there was some volatility in alliances.

We further assume that in a world of alliances, being a member of an alliance is a benefit and, symmetrically, not being a member of an alliance is a cost. Once people can form alliances, individuals left out of the protection of a group are subject to easy exploitation. Evidence that people derive pleasure from membership in groups (Baumeister & Leary, 1995) is indicative of motivational systems executing this function.

In this hypothetical world of shifting group memberships, there would, of course, be many dimensions along which people are evaluated for possible membership in a group. These would presumably have to do with properties of the individual, such as skills, intelligence, physical condition, social connections, and so on.

While these properties are all no doubt important, one key parameter might be the extent to an individual is viewed as likely to change sides as the fault lines of conflict shift. When alliances are dynamic, a member who can, when opportunity arises, shift to the competing group, is extremely dangerous. This suggests that the ability to signal that one will not – or, even better, *can not* – switch alliances can be a

benefit, rather than a cost, because committing can make one a more valuable group member (Frank, 1988). This idea is a specific case of the general notion that removing one's own options can be strategically advantageous if it is signaled to others (Schelling, 1960).

This idea might help to explain various practices surrounding group membership. Scarification – the practice of making permanent marks on one's skin with colors or shallow cuts – might be designed to help persuade others that one is committed to one's group (e.g., Rush, 2005). To the extent that rivals would not accept an individual with these permanent marks into their group, these signals are honest in the technical sense of the term.

Scarification and tattoos (like false beliefs) can be dangerous, leading to the possibility of damage or infection. Despite this, it is still practiced widely, pointing to the possibility of an evolved appetite for visible signals of commitment – whether to groups or romantic partners.

Supernatural Beliefs as Commitment

Beliefs, unlike scars and tattoos, are invisible and easily revised. Spoken statements are themselves ephemeral, limiting their effectiveness as commitment devices. Having said that, giving rise to a belief in another person's head can, under certain circumstances, recruit the power of commitment. For example, as Frank (1988) discusses, information that makes one vulnerable can be useful in this context. If Alfred tells Bob information that would be disastrous for Alfred should it get out, Alfred has, effectively, assured Bob that he won't act in such a way that would make Bob unfavorably disposed towards him. When Bob knows information

that would compromise Alfred – perhaps where to find evidence of a crime that Alfred has committed – Bob can be assured of Alfred's loyalty. So, *transmitting certain kinds of information to others can increase the extent to which they are likely to believe you will remain a loyal ally, which can yield important benefits.*

Broadcasting beliefs might allow commitment. For example, public statements of loyalty to a particular group – or antipathy for other local groups – might help assure potential allies of one's commitment. However, talk is cheap, and such pronouncement do not bind one's actions in the same way that tattoos, scars, or disclosing incrimination information does. Opinions can change, apologies and restitution can be made.

Some statements, however, might make one what Boyer (personal communication) has called "unclubbable," meaning undesirable as a member of a group or community. Such statements, according to the logic of commitment above, are, to be clear, potentially *good things*: From the standpoint of commitment, ways to disqualify oneself from group memberships are the goal.

Consider the following statements:

1. *Columbus discovered America in 1215.

2. *The Earth is flat.

3. *I enjoy eating my own feces.

Statements 1 and 2, in modern times, would, it seems reasonable to say, invite relatively negative evaluations. Everything else equal, people prefer group members who do not have beliefs that are thought to be obviously false. However,

even if it were known that someone had such false beliefs, they would not necessarily be subject to social exclusion.

Statement 3, in contrast, as long as it is not said in obvious jest, would be particularly likely to elicit negative evaluations. As the literature on social stigma suggests, such deviations from normal human behavior elicits very strong negative evaluations (Kurzban & Leary, 2001).

The problem with 1 and 2, then, is that they are not strong enough – they don't make you unclubbable in *any* group. Statement 3, in contrast, is too strong – it makes you unclubbable in *every* group.

So, *to solve the commitment problem, what is required is the sincere endorsement of a belief which makes one unclubbable in every group* except *the group to which one is trying to signal loyalty*. What sort of belief will make one a poor candidate for group membership in nearly every group *except* the one that one is currently in or wishes to commit to?

To return to Statement 3, what makes someone unclubbable about this is the departure from canonical human nature. Human social cognitive systems appear designed to sift through the social world, evaluating others as potential mates, allies, and group members. Departures from the skeletal structure of basic features of human nature act as cues that count heavily against candidates for social interaction (Kurzban & Leary, 2001).

Recall our discussion of Boyer's (1994) ideas surrounding intuitive ontology. To a first approximation, by virtue of shared human computational architecture, people share intuitive ontological commitments. Supernatural beliefs violate these

commitments. In this sense, supernatural beliefs are singularly good at making one appear to have beliefs that violate fundamental causal intuitive principles. In this, they are very different from garden variety false beliefs. Beliefs 1 and 2 are false, but their falseness does not come by virtue of a conflict with intuitive ontology.

In this sense, *supernatural beliefs might be well suited to making one unclubbable because they connote deviation from the species-typical design*. Individuals who do not respect the basic principles that govern causal reasoning about fundamental categories in the world – ARTIFACTS, ANIMALS, and PEOPLE – are, by and large, seen (with a key exception) as mentally ill.

The Diagnostic and Statistical Manual reflects this idea. In the DSM-IV, a "delusion" is defined this way: "A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everybody else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary." Harris (2005) points out the similarity between a SNB and a delusion. "We have names for people who have many beliefs for which there is no rational justification. When their beliefs are extremely common we call them "religious"; otherwise, they are likely to be called "mad," "psychotic," or "delusional" (p. 72).

The key point is that supernatural beliefs will be easily identified by people as false because of people's intuitive ontological commitments. This will lead people to infer that the person who endorses such beliefs – and "firmly sustains" them – is, to a first approximation, insane. The mentally ill are one of the most heavily of stigmatized groups (Corrigan, 2005).

This has one very large exception, as indicated by the definition in the DSM-IV. *False beliefs that that are shared by "almost everybody else" are not considered delusions*. Consider the following.

4. *Eating another person gives you access to their soul.

5. *If a special person says special magic words in a special building, certain crackers turn into the body of a person who was alive but is now dead.

6. *A certain kind of tree can be made to fruit if a pretty woman kicks it.

7. *Keeping your dead grandmother's hair in a jar keeps her spirit around.

First, it is worth asking if one can intuit which of these beliefs are supernatural beliefs culled from the world's cultures and which are delusional beliefs culled from the clinical literature. (There are two of each.[1])

People who endorse such beliefs might be taken for either mentally ill or not, *depending on the social context in which such beliefs are uttered*, specifically whether or not the supernatural belief is commonly held by the other people in a social group. Among those who believe in the transubstantiation, (5) will not make one appear mentally ill. Indeed, endorsing this belief not only does not elicit exclusion, but in fact, in some communities, is essentially a *requirement for inclusion*. Wright (2009) quotes an interesting observation of this general phenomenon suggested by the apostle Paul, who asks, if "the whole church comes together and speaks in tongues, and outsiders or unbelievers enter, will they not say that you are out of your mind?" (p. 270).

The very first commandment, of course, echoes this idea. The call to monotheism, and the harsh punishments in the Old Testament for polytheism, is

consistent with the idea that SNBs are for preventing membership in other groups. The first commandment essentially prevented switching in a world in which other groups were worshiping multiple deities. In this sense, the modern practice of religious tolerance can be seen as evidence for, rather than against our position. The massive efforts that must be made to try to get people to be tolerant of others' religious views suggests that this is not the default state.

Related, Iannaccone (1994), drawing on earlier arguments by Kelley (1986), suggests that religious groups are successful because the things that make them distinctive "invite ridicule, isolation, and persecution," (p. 1182), and that such groups "demand of members some distinctive, stigmatizing behavior that inhibits participation or reduces productivity in alternative contexts…" (p. 1188), ideas that resonate closely with the notion that supernatural beliefs are effective ways to commit to one group over others. Note that Iannacone (1994), however, suggests that the benefit of such costs has to do with public goods, rather than the present argument.) He quotes Singh (1953): "The Guru wanted to raise a body of men who would not be able to deny their faith when questioned, but whose external appearance would invite persecution and breed the courage to resist it" (Singh 1953, p. 31). (See also Iannaccone, 1992.)

Summary

To summarize, our argument begins with the notion that beliefs that preclude membership in other groups are valuable because of commitment. Supernatural beliefs, which violate the basic ontological commitments of evolved intuitive theories, make one appear, to those who do not share such beliefs, mentally

ill, an idea reflected in modern psychiatric classification. If supernatural beliefs do have this property, then there could have been selection for mechanisms designed to generate and endorse locally distinctive supernatural beliefs. Such a mechanism potentially solves the commitment problem by allowing one to preclude membership in any groups other than the local one.

Supernatural beliefs have advantages over other potentially distinctive local beliefs. For example, false beliefs about history, while they might be locally distinctive, do not preclude membership in other groups. Supernatural beliefs, unlike other beliefs that might be locally shared, have the particular property of committing one to the local group that shares the supernatural belief, making them functional in a way that essentially any non-supernatural belief could not. This gives a functional explanation for Boyer's (1994) finding regarding supernatural beliefs, and might help to explain how the costs of false beliefs might be offset.

It seems plausible – though this is not central to the present argument – that rituals might be ways to signal one's endorsement of the false belief that goes beyond simple statements to that effect. Taking communion, for example, might help to persuade others that one endorses (5). Other rituals, instead of being costly signals, might be means of persuading others that one really endorses particular supernatural beliefs. This changes the value of ritual from signaling cost *per se* to signaling belief.

Implications for Intergroup Conflict

One puzzling feature of religious conflict is the degree of antipathy between groups that share nearly all of their supernatural beliefs, with only a handful of such

beliefs distinguishing them. The various antipathies of the world's major monotheistic religions are well known, as is the blood spilt over details of supernatural beliefs among the divisions of Christianity. One might have predicted that similarity reduced hostility, with, say, monotheistic Catholics most fiercely antagonistic toward polytheistic Hindu, but less towards Mormons. This does not, however, seem to be the case. Despite massive overlap in large numbers of false beliefs, a tiny number of such beliefs that differ seem to be sufficient for striking negative emotion and hostility, as one sees in fights among sects.

There are, of course, many possible explanations for this phenomenon, including the fact that groups with similar beliefs might be engaged in conflict for the same resources (Wright, 2009) because of their proximity, but it sits well with the present view. If supernatural beliefs are designed specifically for the purpose of committing people to particular groups because of the potential for conflict, then it is not surprising that such differences should breed fear and hostility.

Along similar lines, the present view resonates well with the fact that organized religions are the locus of trust and cooperation (Wilson, 2002). If shared supernatural beliefs are a good cue to group commitment, then they ought to bring about emotions of trust and support. In the context of intergroup competition, mutually beneficial within-group transactions are very valuable. It is worth noting that there is nothing in and of itself that suggests that false beliefs held in common would lead to trust and strong community ties.

The foregoing suggests that supernatural beliefs should play a special role in both within and between-group social relationships. Within groups, shared

supernatural beliefs, and any acts that are indicative of such shared beliefs (such as particular rituals), should make others feel that the person in question is trustworthy and a loyal member of the group. This should be particularly the case for public activities, which would be serve the function of disqualifying one from membership in other groups. This is distinct from other kinds of beliefs. For example, false shared historical beliefs should not lead to inferences of trustworthiness in the same way that supernatural beliefs might.

In short, we argue that supernatural believes are not, in themselves, accidental consequences of design, nor is the fact that they are at the center of intergroup conflict an accidental consequence of design. On the present view, then, mechanisms that give rise to supernatural beliefs that cause their bearer to be feared and hated by others who do share the belief are functioning precisely as they were designed.

Endnote

[1] Of these, (4) and (7) are drawn from clinical accounts, while (5) and (6) are religious beliefs.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental*

disorders (4th ed., text revision). Washington, D. C.: Author.

Barrett, J., & Nyhof, M. (2001). Spreading nonnatural concepts. *Journal of Cognition and Culture, 1,* 69-100.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, MA: Cambridge University Press.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117,* 497-529.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Boyer, P. (1994). *The naturalness of religious ideas: A cognitive theory of religion.* Berkeley, CA: University of California Press.

Boyer, P., & Bergstrom, B. (2008). Evolutionary perspectives on religion. *Annual Review of Anthropology, 37,* 11-130.

Boyer, P., & Ramble, C. (2001). Cognitive templates for religious concepts: Cross-cultural evidence for recall of counter-intuitive representations. *Cognitive Science, 25*, 535-564.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind*. Chicago: University of Chicago Press.

Corrigan, J. (2005). Is the experimental auction a dynamic market? *Environmental & Resource Economics, 31,* 35-45.

Cosmides, L., & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. In J. Dupré (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 277-306). Cambridge, MA: MIT Press.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7*, 173-179.

Dawkins, R. (1982). *The extended phenotype: The gene as the unit of selection.* Oxford: W.H. Freeman & Co.

Dawkins, R. (2006). *The God delusion*. New York, NY: Bantam Books.

Dennett, D. C. (2006). *Breaking the spell: Religion as a natural phenomenon*. New York, NY: Penguin Books.

Dennett, D. C. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition, 112*, 281-299.

Fodor, J. (2000). *The mind doesn't work that way.* Cambridge, MA: MIT Press.

Frank, R. (1988). *Passions within reason: The strategic role of the emotions.* New York, NY: W.W. Norton & Co.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology, 144,* 517-546.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: John Wiley and Sons.

Harris, S. (2004). *The end of faith: Religion, terror, and the future of reason.* New York, NY: W. W. Norton & Co.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78,* 81-91.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review, 10*, 47-66.

Iannaccone, L. R. (1994). Why strict churches are strong. *American Journal of Sociology 99,* 1180-1211.

Iannaccone, L. R. (1992). Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy, 100*, 271–291.

Irons, W. (2001). Religion as a hard-to-fake sign of commitment. In R. Nesse (Ed.), *Evolution and the capacity for commitment* (pp. 292-309). New York, NY: Russell Sage Foundation.

Johnson, D. D. P., & Bering, J. M. (2006). Hand of God, mind of man: Punishment and cognition in the evolution of cooperation. *Evolutionary Psychology, 4,* 219–233.

Kelley, D. M. (1986). *Why conservative churches are growing: A study in sociology of religion with a new preface for the ROSE edition.* Macon, GA: Mercer University Press.

Kurzban, R. (in press). *Why everyone (else) is a hypocrite: Evolution and the modular mind*. Princeton University Press.

Kurzban, R., & Aktipis, C. A. (2007). Modularity and the social mind: Are psychologists too selfish? *Personality and Social Psychology Review, 11,* 131-149.

Kurzban, R., & Aktipis, C. A. (2006). Modular minds, multiple motives. In A. Shaller, J. Simpson, & D. Kenrick (Eds.), *Evolution and social psychology* (pp. 39-53). New York, NY: Psychology Press.

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin*, *127,* 187-208.

Kurzban, R., & Neuberg, S. (2005). Managing ingroup and outgroup relationships. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 653-675). Hoboken, NJ: Wiley.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98*, 15387-15392.

McKay, R. T., & Dennett, D. C. (in press). *The evolution of misbelief.* Behavior and Brain Sciences.

Milikan, R. (1984). *Language, thought, and other biological categories.* Cambridge, MA: MIT Press.

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle*. Evolution and Human Behavior, 26,* 88-105.

Nesse, R. M. (2001). *Evolution and the capacity for commitment.* New York, NY: Russell Sage Foundation.

Nesse, R. M., & Lloyd, A. T. (1992). The evolution of psychodynamic mechanisms. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 601-624). New York, NY: Oxford University Press.

Nesse, R. M., & Williams, G. C. (1994). *Why we get sick: The new science of Darwinian medicine*. New York, NY: Vintage Books.

Nettle, D. (2004). Adaptive illusions: Optimism, control and human rationality. In D. Evans & P. Cruse (Eds.), *Emotion, evolution and rationality* (pp. 193–208). Oxford, England: Oxford University Press.

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York, NY: HarperCollins.

Rush, J. A. (2005). *Spiritual tattoo: A cultural history of tattooing, piercing, scarification, branding, and implants*. Berkeley, CA: Frog, Ltd.

Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Sidanius, J., & Kurzban, R. (2003). Evolutionary approaches to political psychology. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Handbook of political psychology* (pp. 146-181). Oxford: Oxford University Press.

Singh, K. (1953). *The Sikhs*. London: Allen & Unwin.

Smith, M. J. (1982). *Evolution and the theory of games.* Cambridge, MA: Cambridge University Press.

Sosis, R., & Alcorta, C. (2003). Signaling, solidarity, and the sacred: The evolution of religious behavior. *Evolutionary Anthropology, 12*, 264–274.

Sperber, D. (1985). *On anthropological knowledge*. Cambridge, MA: Cambridge University Press.

Tiger, L. (1969). *Men in groups.* New York, NY: Random House.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 19-136). New York, NY: Oxford University Press.

Trivers, R. (2000) .The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences, 907,* 114-131.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wiley, H. R. (1994). Errors, exaggeration, and deception in animal communication. In L. Real (Ed.), *Behavioral mechanisms in evolutionary ecology* (pp. 157-189). Chicago: University of Chicago Press.

Wilson, D. S. (2002). *Darwin's cathedral: Evolution, religion, and the nature of society.* Chicago: University of Chicago Press.

Wright, R. (2009). *The evolution of God.* New York, NY: Little, Brown and Company Hachette Book Group.

Zahavi, A. (1975). Mate selection: A selection for a handicap. *Journal of Theoretical Biology, 53,* 205-214.