

Oxford Handbooks Online

Supernatural Beliefs and the Evolution of Cooperation



Pierrick Bourrat and Hugo Viciano

The Oxford Handbook of Evolutionary Psychology and Religion

Edited by Todd K. Shackelford and James R. Liddle

Subject: Psychology, Personality and Social Psychology,
Cognitive Psychology

Online Publication Date: Jun 2016 DOI: 10.1093/oxfordhb/9780199397747.013.23

Abstract and Keywords

Studies have found an association between the content of beliefs in the supernatural and increased cooperation in social groups. “High Moralizing Gods,” “fear of supernatural punishment,” and “supernatural monitoring” have been claimed to permit greater social cohesion through the specific epistemic engagement they produce in the minds of those who hold certain religious beliefs. However, the evolutionary pathways linking these religious features with cooperation remain unclear. Focusing on the example of belief in supernatural sanctioning, this chapter delineates different mechanisms by which beliefs in supernatural entities could, in principle, lead to greater cohesion and emphasizes the different predictions each evolutionary mechanism affords. It thus reassesses several studies that have been interpreted as supporting or as failing to support one or some of these cultural evolutionary processes. Finally, it proposes several avenues by which research addressing the link between cooperation and specific forms of belief in supernatural entities could be strengthened.

Keywords: prosociality, supernatural punishment, religious, belief, vigilance, adaptationism, cooperation, Big God, High God

Background: New Wine in Old Bottles?

Philosophical debate regarding the utility of specific religious beliefs has existed for almost as long as philosophy itself. Whereas Plato (427–348 BC) thought that certain false supernatural myths could anchor social harmony in the perfect republic, other philosophers such as Blaise Pascal (1629–1662) concluded that belief in the Christian

God was a good bet for the individual, even if the belief happened to be false—an idea known as Pascal's Wager. During the Enlightenment, philosophers such as Baron D'Holbach and Denis Diderot rejected the "false religion" that injected fear in the masses for the exclusive benefit of priesthood while pleading for a decoupling of morality and particular religions. On the contrary, other public intellectuals such as Edmund Burke (1729–1797) considered morality and a specific form of religious belief inseparable.

With the advent of scientific anthropology and sociology, opposing sides also formed. In one camp, those such as Emile Durkheim (1858–1917) emphasized the social utility of the more external, ritualistic components of religious forms, whereas others such as Edward Tylor (1832–1917) highlighted what, in their view, was a weak link between morality and religion in most preagricultural societies, affirming that even if certain specific contents of religious beliefs encouraged morality, other beliefs such as those associated with ancient animism were mostly "unmoral" (Tylor, 1871/2010, p. 326).

It was only during the second half of the 20th century that this issue was addressed through more systematic means and empirical research. Combined with an evolutionary approach, empirical research allows for a detailed formulation of what are, in principle, falsifiable hypotheses. An adaptationist methodology (Andrews, Gangestad, & Matthews, 2002) can be put to use to confirm or disconfirm functionalist accounts. The positive effects of holding certain beliefs on the spread of these ideas must then be scrutinized. Such corroboration or falsification of hypotheses can only be attained by assessing a number of possible mechanisms of diffusion and maintenance in the population. Evolutionary psychology offers a framework for formulating hypotheses as well as new tools for testing them (Pinker, 2011). This research strategy has already achieved some progress in understanding the social effects of beliefs in the supernatural. However, as we shall see, important aspects may need further work requiring the use of innovative methods.

Evolutionary Accounts of Religious Belief Based on Cooperation

Belief in some form of supernatural entities is a cultural phenomenon found in every human society (Benedict, 1938; Brown, 1991; Murdock, 1945), as are moralistically enforced norms and rules (Brown, 1991, p. 139; Turiel, 2002). In recent years, a body of work has attempted to provide an evolutionary account of the origins and maintenance of these phenomena. Whereas a prominent account on the evolution of religion considers that beliefs in supernatural agents are mainly a byproduct of human cognition, others have presented hypotheses in which such beliefs are regarded as some form of adaptation, at the level of the individual or the group (Bourrat, 2015c; Viciano & Bourrat,

2011). The role of “supernatural monitoring,” “supernatural punishment,” “Big Gods,” and “High Moralizing Gods” in the evolution of cooperation has received particular attention (Atkinson & Bourrat, 2011; Bourrat, 2015c; Johnson & Krüger, 2004; Norenzayan, 2013; Shariff & Norenzayan, 2007; Watts et al., 2015). Several studies claim that the greater prevalence or activation of these beliefs may predict greater cooperation. However, several questions remain. Are certain forms of beliefs in the supernatural an adaptation linked to increased cooperation? Or are they epiphenomenally linked with cooperation (i.e., their specific content has no meaningful causal impact on social cohesion and the enforcement of morality)? And how can a cultural evolutionary perspective inform research on this topic?

The anthropologist Evans-Pritchard et al. (1954; quoted in Swanson, 1964, p. viii) wrote:

Generalizations about “religion” are discreditable. They are always too ambitious and take account of only a few of the facts. [...] Sweeping generalizations reached by dialectical analysis of concepts ... [should be] abandoned if favour of limited conclusions reached by inductive analysis of observed facts.

Here we want to depart slightly from this type of admonishment by considering the details of some potentially valuable theoretical “generalizations.” It is worth looking at these generalizations in the context of the evidence presented in a series of empirical studies. To complete Evans-Pritchard’s statement, one could argue, paraphrasing Immanuel Kant, that theoretical generalizations without ethnographic and sociological analysis are empty, but ethnographic analysis without theoretical generalizations is blind. Evolutionary cultural adaptationism can play an important role in adequately specifying and assessing a number of hypotheses.

Importantly, although there are several mechanisms by which beliefs in supernatural entities could increase cooperation, these mechanisms are not clearly differentiated in the literature (for a notable exception, see Schloss & Murray, 2011). This is problematic because, once differentiated, these mechanisms lead to unique predictions regarding the actual processes at play. As a result, not distinguishing these processes obscures the underlying causal and evolutionary mechanisms operating between beliefs in the supernatural and cooperation.¹

In this chapter, we distinguish different mechanisms by which beliefs in supernatural entities could have increased cooperation and been selectively advantageous. We review some of the evidence for and against hypotheses about supernatural beliefs as causally responsible for increased cooperation. We analyze these results with the theoretical underpinnings of adaptive evolution, but also with alternative hypotheses. Rather than presenting an exhaustive overview of the literature, we review a few examples mostly

focusing on the exploration of the fear of supernatural punishment hypothesis in relation to cooperation. This helps illustrate why the theoretical distinctions we propose are important. Finally, we suggest a few avenues that, if pursued, would permit a better understanding of the links between cooperation and specific beliefs in the supernatural.

Terminology

If we are to consider the hypothesis that religious beliefs evolve as a result of their effect(s) on cooperation, we want to be precise about the nature of the hypothesis. Here we focus on the different hypotheses surrounding supernatural sanctioning and increased cooperation. Thus, before entering into the heart of the matter, we need working definitions of the terms “belief,” “supernatural punishment,” “supernatural agent,” “god,” “Big God,” and “High Moralizing God,” because each one of these concepts has been used in work evaluating the link between supernatural belief and cooperation. There is overlap among these terms, which is one reason why some authors have seemingly used some of them interchangeably.² We make these distinctions for the practical purpose of using terms consistently and as a point of anchor throughout the chapter. Some might disagree with the particular categorization outlined in this chapter. What is most important here, however, is consistency; only through the consistent use of concepts can hypotheses be clearly constructed in terms of falsifiability. We begin with the concept of “belief.”

“Belief” is both a widely used concept and also a generally poorly defined one (Kurzban, 2011). It is sometimes noted that the usual understanding of belief is full of philosophical paradoxes. What one person really believes can be difficult to pinpoint if one has too loose a concept of “belief,” not only because one person’s beliefs can differ from what this person openly claims (as in the phenomena of duplicity, excessive politeness, or lying; see Kuran, 1997) but also because it may not be obvious what one’s beliefs imply. For instance, does a belief in numbers and common rules of addition and subtraction specifically imply that one believes 327 minus 52 equals 275? If Lois believes that Superman has rescued her, does she believe then that Clark Kent is a superhero?

Beliefs, however, are normally considered to have content—that is, they reflect the understanding of *only a specific aspect* (and not others) in the perception of a given reality. The “morning star” and the “evening star” describe different aspects of the same object, namely Venus, even though some ancient Greeks still *believed* them to be different objects. According to recent research in developmental psychology, children can minimally understand this “aspectuality” associated with others’ beliefs (also called *intensionality*) at least as early as 4–5 years old (Rakoczy, Bergfeld, Schwarz, & Fizke,

2015). Nevertheless, the existence of multiple cognitive phenomena that can be associated with the psychological concept of belief suggests that there is not one single reality that corresponds to a belief (Gendler, 2008). For the purposes of this chapter, we consider a belief to be any mental representation or brain process that can be ascribed some particular content or aspectuality in relation to other cognitive processes such as memory, communication, or decision-making.³

Regarding the term “entities,” we mean here both agents (e.g., God) and nonagential objects or concepts (e.g., Hell). Concerning the notion of “supernatural punishment,” we consider in this chapter that it represents a punishment that results from the causal powers of an entity that is nonobservable. For example, going to hell because one does not comply with some moral doctrine is considered a supernatural punishment. One subset of supernatural punishments comprises supernatural punishments professed by supernatural agents such as gods, spirits, and ancestors, who can be defined following Boyer (2001, p. 6) as nonobservable⁴ agents with some ascribed causal powers. Closely related to the notion of supernatural punishment is the notion of supernatural monitoring. We consider that supernatural monitoring represents the monitoring of one’s behavior or thoughts by at least one supernatural agent. The notion of “supernatural agents” is broad and includes ghosts, gods, Big God, and High God. Ancestors, spirits, or Santa Claus are all supernatural agents, and believing in them might *prima facie* be the result of some evolutionary process one might attempt to explain. In his 2013 book, Ara Norenzayan gives an important place to what he calls Big Gods. By contrast to what one might call “minor gods,” Big Gods are, according to Norenzayan (2013, pp. 7–8), “powerful, omniscient, interventionist, morally concerned gods.” Attributing these properties to Big Gods implies that such gods could not be easily fooled or mocked, as some more minor gods are in different societies. Their assumed “bargaining power” is thus stronger than that of minor deities. It should be noted that the notion of a Big God corresponds roughly to the notion of a High Moralizing God, which is the term used in the *Ethnographic Atlas* (EA) (Murdock, 1967a) and the Standard Cross-Cultural Sample (SCCS) (Murdock & White, 1969), on which many of the claims linking cooperation and beliefs in supernatural agents rest (see below, Section 4). In the EA and the SCCS, a High Moralizing God is defined as a type of High God, that is, “a spiritual being who is believed to have created all reality and/or to be its ultimate governor, even though his/her sole act was to create other spirits who, in turn, created or control the natural world” (Murdock, 1967b, p. 160), and is specifically supportive of human morality (Roes & Raymond, 2003). Most predictions about the role of beliefs in supernatural agents, however, do not explicitly link the fact of believing in one single creator of the natural world and human cooperation.

Searching for the Ethnographic “Big Picture”

Although many (perhaps most) in the scientific study of religion have followed Durkheim (1912/1979) in putting special emphasis on the more ritualistic aspects of religious forms (e.g., Sosis & Kiper, 2014), the content of specific religious beliefs has also been intensely studied in relation to its effects on social cooperation. An important body of evidence has been produced in this domain.

Perhaps the most influential and pioneering work in this respect was that of Guy Swanson. Swanson’s *The Birth of the Gods* (1960) sought to corroborate a number of hypotheses on the spread and nature of beliefs in the supernatural with data from the wide ethnographic record as presented in one of the first versions of Murdock’s (1957) “World Ethnographic Sample.” Contrary to Edward Tylor’s earlier stance (1871), Swanson found that the supernatural and moral domains were not entirely unconnected in small-scale societies. However, perhaps his most enduring result (also replicated a number of times: Davis, 1971; Peregrine, 1996; Sanderson & Roberts, 2008; Stark, 2001; Underhill, 1975) was the finding that “High Gods” were disproportionately found in societies with three or more layers of hierarchical structure in terms of distinct “sovereign groups” (e.g., the household, the village, the tribe). Of these societies, however, only a fraction of them presented High Gods that were also moralizing gods (some of these gods being relatively uninterested in humans).

Are religious beliefs, however, necessarily connected with moral sanctioning, even when disconnected from the idea of a High God? The short answer is “no.” Christopher Boehm (2008) conducted an analysis of 43 ethnographies that index the behavior of 18 hunter-gatherer societies specifically selected to be relatively representative of the Late Pleistocene “way of life.” He found that supernatural sanctioning of actions often regarded as immoral behavior (e.g., lying, cheating, stealing, murder) was present in all 18 societies. However, none of these behavioral categories was condemned by supernatural means in *all* 18 societies. In fact, only “incest” was supernaturally sanctioned in at least half of these societies, closely followed by “murder,” which was condemned by supernatural sanction in 8 of the 18 societies. In striking opposition to this, 13 societies had some form of supernatural sanctioning related to food. Other forms of supernatural sanctioning of taboos and rituals had similar frequencies.

High Moralizing Gods, concerned with human social life in general, have been linked to a series of social and ecological characteristics in a number of correlational studies. Studies searching for the ecological factors affecting the differential survival of religious beliefs generally work around the hypothesis that certain religious beliefs may positively

affect the way that populations cope with environmental stress. This creates a feedback loop by way of which the belief persists in the population. They follow the work of John Snarey (1996), who found that among societies with coded data in the SCCS, those located in environments of water scarcity were more likely to profess beliefs in a supreme deity concerned with moral wrongdoing.

More recently, Botero et al. (2014) modeled the effect of a broader range of ecological variables on the presence or absence of belief in moralizing gods in different societies while replicating the main finding by Snarey. Among those studies searching for social determinants, Roes and Raymond (2003) found a statistically significant 0.29 Kendall's τ correlation between society size and presence of belief in High Moralizing Gods among the societies included in the SCCS (but see the section "Cognitive Causal Mechanisms" below) for more on the variable "number of jurisdictional hierarchies beyond the local community" mentioned earlier and which they used as a proxy for society size). Dominic Johnson (2005), after controlling for world region and type of religion and applying Bonferroni corrections for multiple testing, analyzed the SCCS and found that the number of jurisdictional hierarchy levels beyond the community level and the lending of money were positively correlated with the presence of a High Moralizing God in a given society. Interestingly, he did not find a positive correlation with the variable "compliance with society norms," although this variable has been deemed unreliable. Neither was there a statistically significant relationship with the variable "loyalty to the local community." Using the SCCS, Bourrat, Atkinson & Dunbar (2011) failed to find correlations between belief in broad supernatural punishment (where they included belief in High Moralizing Gods, but also other variables, such as belief in witches, or the evil-eye) and ethnographically recorded measures of high levels of individual norm compliance (see the section "Cognitive Causal Mechanisms" for a distinction between individual cooperation and collective-level cooperation).

Others have claimed to find stronger support for the idea that a certain type of religious belief is specifically connected to the moral order. Rodney Stark (2001), using data from the 1990–1991 wave of the World Values Survey (WVS) and his own theological analysis of the importance of a personal God in different contemporary world cultures, found that in those societies in which the importance of a personal God (as opposed to an impersonal force or other theological constructs) is more present, those respondents that affirmed that God was more important in their lives tended to condemn more strongly certain actions, such as buying a stolen good, failing to report damage caused to a parked car, or ingesting marijuana. Interestingly, he found that the self-reported importance of God in one's life was a more predictive factor than church attendance (which generally failed to reach statistical significance). Atkinson and Bourrat (2011) found a similar, stronger pattern when looking at the aggregate data of five waves of the WVS across 87

countries ($n = 355,298$). In their study, beliefs in supernatural concepts such as Heaven, Hell, and a personal God (as opposed to an impersonal spirit or life force) were associated with stronger moral condemnation of actions. The correlation remained significant after controlling for region, level of education, type of religion, and frequency of attendance at religious services.⁵

These studies, however, are correlational and, for the most part, confront one famous difficulty in the study of human history—the so-called Galton’s problem (Naroll, 1965). Francis Galton highlighted that it is not possible to infer ecological causality from observed cultural patterns if one does not first exclude the influence of a simpler form of cultural diffusion in explaining these patterns.⁶ Whereas some of the authors of the aforementioned studies controlled for world region in their statistical modeling, this is still a crude proxy for absence of cultural contact. In fact, the epistemological challenge that Galton’s problem presents—namely, excluding all possibility of contact as a previous stage to test the model—may be impossible to satisfy in all but a few cases. A recent study in Austronesia introduced Bayesian phylogenetic methods to control for patterns in the direction of the appearance of moralistic supernatural sanctioning (Watts et al., 2015). They found that one of the reasons why complex societies have High Moralizing Gods more often may be because they have had more chances to enter into exchanges of their surplus with monotheistic Muslim societies, from which they would have borrowed their concept of a High God through a process analogous to what linguists call “calquing.”

Studies using self-report data secured in large social surveys confront the additional problem of basing what they take to be a measure of prosociality from self-report. However, it is an important point that can be defended, both on conceptual and empirical grounds, that it is not the same to condemn one behavior and to act morally. In other words, moral action and moral judgment are two partially independent phenomena. Moral hypocrisy is a phenomenon that has been shown to be relatively prevalent in a number of famous psychological studies (see Kurzban, 2011). Furthermore, the fact that in certain (mostly Western) societies those individuals who tend to acknowledge that God is more important in their lives also tend to more strongly condemn certain antisocial actions suggests a social desirability bias—a bias which, incidentally, has been found to be prevalent among religious people (Sedikides & Gebauer, 2009).

Evolutionary Mechanisms

Polybius, the second century BC Greek observer, highly valued the attention the Roman community devoted to its dead. The honor descendants publicly devoted to their most famous ancestors in funeral processions pressured them to measure up to the progenitors' example or else be considered unworthy. "Secret fears" (hadelois phobois) and respect for punishment experienced in or emanating from Hades, where the manes dwell, Polybius added, restrained the magistrates. Considering the civic virtue evident in Roman government, Polybius thought, the Greeks had erred in scuttling their religion.

(Bernstein & Katz, 2010, p. 210)

The Greek historian Polybius considered Roman religious beliefs to be an important part of the explanation of the success of their institutions. Elaborating a plausible account of how beliefs may persist in a population as a result of their social effects requires specifying these mechanisms, including some form of feedback loops. How can cultural beliefs linked to the supernatural evolve to sustain some form of increased cooperation? Having settled the ground by showing part of the big picture linking specific beliefs in the supernatural and human morality, we now review different processes or types of evolutionary mechanisms, with different degrees of plausibility, by which certain specific beliefs in the supernatural could lead to different forms of cooperation.

By cooperation, we mean here any social behavior that can be exploited by another individual (i.e., is subject to free-riding or cheating). Classically, if a cooperator pays a cost c to contribute to the production of a good b shared by the members of a community (often including the focal individual), it becomes tempting and evolutionarily advantageous to receive b without having to pay c (Bourrat, 2015d). It is thus expected that in a single community, whenever a situation of that sort arises, cheaters or "free-riders" will invade the population, which often results in overexploiting the resources and leads to a tragedy of the commons (Hardin, 1968). Thus, without means to stop cheaters from invading the population, cooperation at the community level will not evolve. Cheaters can be stopped from invading the population either through adaptations or byproducts of other evolutionary processes (Schloss & Murray, 2011). In any case, the means to stop cheaters from invading are sine qua non conditions for cooperation to evolve.

When the population is structured such that individuals can only interact with their neighbors, it is expected that the cheater strategy will not always invade the population. A comparison between the spread of cheating and the spread of viruses may be of use here. Consider the following evolutionary mechanism by which the decrease in virulence can be explained. If a pathogen transmitted from one individual to the other exploits its host quickly and thus gains a faster growth rate, this strategy will be profitable as long as

there is a sufficiently large supply of hosts (Bull, 1994). When hosts become scarce, it becomes evolutionarily advantageous to pay the cost of not reproducing and preserving the host in order to gain the benefit of having more hosts to colonize later. In what follows, we explore the space of possibilities for mechanisms that, analogously, have an effect of restraining cheaters from invading the population through certain beliefs in the supernatural.

The Crudest Hypothesis

Let us start simply. Perhaps the crudest hypothesis linking supernatural beliefs and cooperation is the view that beliefs in supernatural entities can prevent individuals from defecting if it is believed that defecting will be punished, thus inflicting a cost on the individual that defects (see Table 1). Since punishing is costly and therefore subject to second-order defection, transferring the cost of punishing to a supernatural entity arguably solves this problem.

Obviously, as an evolutionary argument, something is lacking; this mechanism per se does not prevent first-order defection from evolving. Suppose a population in which everyone is motivated to pay c by the threat of punishment by a supernatural entity. But a new variant (a “mutant” in terms of evolutionary dynamics) arises who is not afraid to be punished. Because this individual does not pay c , its fitness is superior to other noncheating individuals. As a consequence, we can predict that it will invade the population and disrupt any large-scale social cohesion. In other words, even if the supernatural punishment hypothesis were theologically sound, it still would lack something to constitute a strategy that can evolve and sustain cooperation.

By itself, the fear of supernatural punishment hypothesis is insufficient to explain why belief in supernatural entities would be associated with higher levels of cooperation, because it does not represent a solution to first-order cheating. To be considered a serious contender, it must be supplemented. One way is to suppose, as is frequently observed, that the punishment will occur in the afterlife (e.g., going to Hell for eternity or being reincarnated as a nondesirable creature). Although this solution seems appealing at first, it is fallacious. Of course, if a selection process is at work, any delayed and often nonobservable benefit or cost will have no causal effect on the selective process at work, especially if the benefit is to come after the life cycle of the individual (i.e., the “afterlife”). Again, imagine one mutant that did not believe they would be punished in the afterlife and as a result would act accordingly and not pay c . This mutant would thrive and invade the population, without being punished in the material world. Thus, cooperation under this model is not stable and does not represent a solution to first-order cheating. One of the earliest arguments of that sort can be found in the writings of Pascal

in what has been called Pascal's wager (Hacking, 2001), as mentioned earlier. Pascal's wager does not represent, per se, an evolutionarily stable strategy.

Byproduct Hypotheses

A more sophisticated form of the fear of supernatural punishment hypothesis is to suppose that by lacking belief in supernatural punishment, an individual would necessarily incur some other cost. Such loss would be costlier than the benefit earned by not paying c , which is the cost of restraining from reaping the benefits because of the fear of supernatural punishment. For instance, if we suppose that the belief in supernatural punishment is one effect of a cognitive process among other effects (something like a byproduct) and that not holding this belief would necessarily involve other changes that would have negative fitness consequences for those individuals, then fear of supernatural punishment could represent a solution to first-order cheating.⁷ This would be a result of constraints imposing costs on those deviating from the specific belief in the supernatural. Thus, under this model, first-order cheating with respect to cooperation would decrease fitness overall.

Of course, one important question in regard to this type of hypothesis is whether beliefs in supernatural punishment can be byproducts of cognitive processes that have evolved for other purposes. In addition, it seems unlikely that among all the beliefs in supernatural punishment one individual can have, each one would be the result of cognitive processes that would make them impossible to forsake without disrupting their underlying cognitive processes (with overall negative individual fitness consequences). In other words, it is unlikely that each belief in supernatural punishment is the byproduct of other vital cognitive processes. However, that could be the case for some of them, at least in principle. Suppose for instance that the fear of supernatural punishment is, in some societies, part of the normal cognitive development of children, almost in a similar way as language. As a result, unlearning this particular instance of fear of supernatural punishment would be nearly impossible in the same way that unlearning a mother tongue is impossible. If this hypothesis were to be verified, it could explain why cooperation is facilitated in large-scale societies by beliefs in supernatural punishments. It would also predict, for instance, that people who started to believe in a given supernatural punishment when they were young and people who started to hold this belief as adults may not pay the same cost in forsaking the belief. This would be because acquiring the belief young would constrain adult psychology to a greater extent than acquiring the belief later.

Byproduct hypotheses abound, and it may be challenging to disentangle them from purely adaptationist hypotheses. Another potential cognitive process that could satisfy a

byproduct hypothesis, in the case of the belief in supernatural punishment, is to invoke the causal role played by other cognitive processes through which humans develop a sense of fairness, and which can lead them to hold certain intuitive beliefs related to supernatural punishment. Such is the case, for instance, of the intuitive beliefs about immanent justice that may be easily triggered when a person commits a morally wrong action (see Baumard & Chevallier, 2012). Such a sense of fairness or moral balance could have evolved to prevent individuals from being exploited (for an elaboration of this idea, see Baumard, André, & Sperber, 2013). However, once in place, it is possible to imagine that after observing somebody committing a misdeed, beliefs in supernatural punishment may have become a salient idea (and one that one may not be interested in running the risk of disproving).

These hypotheses can find their root in the so-called byproduct theory of religion first put forward in its modern form by cognitive anthropologists such as Pascal Boyer (e.g., Boyer, 2001) and cognitive psychologists such as Justin Barrett (e.g., Barrett, 2004). We refer to this as the “classical byproduct theory of religion.” The classical byproduct theory of religion is based on the assertion that humans often come to believe in supernatural agents because of hypersensitive agency detection, which results from the asymmetric fitness cost of missing an agent in their environment (potentially very high cost) compared with believing there is an agent in the environment when there is none (quite low cost). Further, as the theory states, humans believe that supernatural agents care about human affairs because humans have a “theory of mind” (TOM) module, which assigns intentions and goals to the agents they interact with (whether supernatural or not). In the same way as missing an agent in the environment when there is one could be extremely costly, failing to attribute certain intentions to the agents one interacts with would represent a high evolutionary cost, higher than the cost of attributing intentions to agents that do not exist. Following this theory, beliefs in supernatural agents that have intentions represent nothing more than a predictable byproduct of human cognitive processes.

It should be noted that the classical byproduct theory of religion and the byproduct hypotheses with respect to fearing supernatural punishment presented in the previous paragraphs need not oppose each other. It is compatible to suppose that, due to the TOM module and the hypersensitive agency detection devices (HADDs) of the brain, individuals detect agency and give them intentions in the environment (classical byproduct theory) and that, because of their intuitive sense of fairness, for instance, they come to fear being punished for misdeeds. However, these hypotheses are not equal in their predictions with regard to cooperation. In fact, the classical byproduct theory of religion does not predict that certain beliefs in the supernatural increase cooperation (Baumard & Boyer, 2013). On the other hand, under the byproduct hypothesis established on a preexisting evolved

moral sense or other cognitive constraints, cooperation could, at least in principle, come for free, because fearing the supernatural entity causally responsible for reestablishing justice may result in some forms of increased cooperation.⁸

The putative evolutionary processes we have presented here, namely the developmental-constraint hypothesis and the sense-of-fairness hypothesis, represent only two possible cognitive processes by which cheating may be constrained by specific beliefs in supernatural punishment (by some agent or others), but there may be many other mechanisms involved. The common feature of all these mechanisms would be that the fear of supernatural punishment that transforms the short-term (or local) fitness benefit gained by cheaters into a long-term fitness cost would result from a byproduct of other cognitive mechanisms. The fear of supernatural punishment, under this generic hypothesis, is a first-order byproduct of cognitive processes evolved for other purposes, and cooperation is a second-order byproduct of the primary byproducts. We refer to this generic hypothesis as the “double-byproduct fear of supernatural punishment hypothesis” (see Table 1). Note, however, that ecological and cultural differences should be invoked to account for why the belief in supernatural punishment should be more important for cooperation in some societies than in others.

Adaptationist Hypotheses

Johnson and Bering (2006) have put forward another hypothesis regarding the evolution of belief in supernatural punishment; instead of regarding the cognitive mechanisms that lead to cooperation as byproducts of other cognitive processes, their hypothesis supposes that belief in supernatural punishment is an adaptation at the individual level with the function of managing reputation. The reasoning underlying this hypothesis is as follows: With the emergence of human language, it became possible to learn information of the kind “who did what to whom” without directly observing it. This in turn led to the emergence of reputation. Reputation has important consequences for human fitness. Thus, the hypothesis is that a lower threshold for believing in the possibility of supernatural punishment could be co-opted by natural selection and prevent cheating behaviors, which if observed by other members of the group could lead to a lower reputation and, consequently, lower fitness for the individual.

Although this hypothesis is different from the double-byproduct fear of supernatural punishment hypothesis, it can also be reformulated more generally in terms of costs and benefits on fitness. Whereas under the double-byproduct fear of supernatural punishment hypothesis, the short-term benefit of cheating is constrained by fearing supernatural punishment due to some cognitive constraints that evolved for different purposes, in this case it is constrained by a cognitive mechanism that is an adaptation. Without this

adaptation, individuals would incur a much greater fitness cost in the long term (through a lowered reputation). Of course, it would be difficult to separate empirically the double-byproduct fear of supernatural punishment hypothesis from the latter hypothesis one might want to call “one byproduct–one adaptation fear of supernatural punishment hypothesis” (see Table 1) because they make largely the same predictions.⁹

Yet another closely related hypothesis should be added—namely, the implicit monitoring hypothesis. Under this view, beliefs in supernatural agents do not elicit a “fear” in supernatural punishment, but rather “hack” a cognitive module that tracks whether one is being watched. There is a growing body of evidence suggesting that being watched makes individuals more cooperative (e.g., Bateson, Nettle, & Roberts, 2006; Bourrat, Baumard & McKay., 2011; Haley & Fessler, 2005; Mifune, Hashimoto, & Yamagishi, 2010). The idea underlying supernatural monitoring is similar to that of the fear of supernatural punishment hypotheses. There is again an adaptationist account and a byproduct account. Beliefs in supernatural agents could, as a byproduct, make individuals feel watched more often and thus lead them to cooperate more (double-byproduct hypothesis). Or, being watched performing an uncooperative behavior could have, on average, deleterious consequences on reputation and thus fitness. Hence feeling monitored supernaturally by agents could be an adaptive strategy (one byproduct–one adaptation hypothesis). Again, it would be hard to distinguish them empirically.¹⁰ Furthermore, there is the difficulty of empirically distinguishing fear of supernatural punishment hypotheses involving specific beliefs about the type of agent (e.g., High Moralizing Gods) from simple supernatural monitoring. To do so would require one to contrast the effects of beliefs on supernatural agents that inspire fears and supernatural agents that do not.

A final hypothesis linking supernatural beliefs and cooperation states that beliefs in supernatural punishment increase within-group cooperation in competition between cultural groups (see Table 1). Competition can be understood as competition for access to resources or, more generally, as a struggle for persistence. It is possible to conceive that different beliefs in supernatural entities by the members of a cultural group lead to different outcomes at that level. Under this hypothesis, the reason why an individual of a particular culture would believe in supernatural punishment is because it would have been transmitted as a result of having been advantageous for their cultural group in the past (the belief would have allowed the group to persist longer in competition with other cultural groups). Persistence of a cultural group can be understood in two different ways. It can mean persistence of the biological individuals having the beliefs or merely the persistence of the cultural traits themselves. In the latter case, cultural group selection could occur even without the death or reproduction of a single biological individual. By persistence of a cultural group, unless stated otherwise, we mean the former case, in

which beliefs are tied to biological fitness.¹¹ Of course, at a purely cultural level, some “ideas” or beliefs are more likely to “survive” in the mind of their “hosts” and can be organized in “groups” and transmitted from one individual to another within the group and even beyond. But it is not always clear what sort of cultural entities could serve as a basis by which to measure fitness at that level.

Following the cultural group hypothesis of fear of supernatural punishment, beliefs in some particular supernatural entities can have a lasting impact on within-group cooperation and prevent groups from dissolving, and can increase within-group cooperation and intergroup hostility. Although this hypothesis considers cultural groups to be the unit of selection, it should be regarded as an alternative way to explain the evolution of cooperation in which the time scale of the events occurring between groups is typically hard to represent from the point of view of individuals (see Bourrat, 2014, 2015a, 2015b). Ultimately, everything could be represented from the point of view of individual fitness, in which case the cost of cooperation would not be represented as lifetime costs and benefits but rather over many generations. Thus, for example, a cost could be paid by an individual at generation N and lead to a benefit at generation $N + 3$ —that is, received by the descendants of this individual. In classical models of cooperation, costs and benefits are represented over the lifetime of an individual (e.g., West, Griffin, & Gardner, 2007). This hypothesis is compatible with all the other hypotheses mentioned in this section. What is regarded as a byproduct in the short term could produce a selective advantage when considering long-term effects in settings with multiple cultural groups.

Table 1 Different Evolutionary Hypotheses Linking Supernatural Sanctioning and Cooperation

Hypothesis	Supernatural Beliefs	Cognitive Processes Potentially Involved	Type of Evolutionary Explanation
Crude supernatural punishment	Belief in supernatural entities (agents, concepts, objects)		Not an evolutionarily stable strategy
Double-byproduct fear of supernatural punishment	Belief in supernatural entities (agents, concepts, objects)	Hypersensitive agency detection device (HADD) and theory of mind (TOM) in the case of beliefs in supernatural agents; Sense of fairness (?) Any cognitive process that potentially can make the cost of cooperation incompressible, such as a developmental constraint	Byproduct at the individual level Byproduct at the group level
One byproduct-one adaptation fear of supernatural punishment	Belief in supernatural entities (agents, concepts, objects)	HAAD and TOM in the case of beliefs in supernatural agents; Cognitive process managing reputation	Partially byproduct and partially adaptive at the individual level Byproduct at the group level

Double-byproduct supernatural monitoring	Belief in supernatural agent	HAAD and TOM	Byproduct at the individual level B-product at the group level
One byproduct-one adaptation supernatural monitoring	Belief in supernatural agent	HAAD and TOM	Partially byproduct and partially adaptive at the individual level Byproduct at the group level
Cultural group-level fear of supernatural punishment	Belief in supernatural entities (agents, concepts, objects)	All of the above	Byproduct or adaptive at the individual level (but difficult to explain) Adaptive at the group level

Before moving on to the next section, it should be noted that none of the evolutionary hypotheses provided in this section emphasize the role for High Moralizing Gods or Big Gods in the sense that fear of punishment could be inspired potentially by *any* supernatural entity and monitoring by *any* agent. Furthermore, it should be stressed that several of the hypotheses presented specifically focus on supernatural *agents* as opposed to simply supernatural *entities*. Many beliefs in supernatural entities could have a similar effect as beliefs in supernatural agents on cooperation. Some hypotheses, like the supernatural monitoring hypotheses, specifically predict that supernatural *agency* is a distinctive feature that could permit increased cooperation between the members of a group, but there is nothing in the hypotheses that predicts that particular supernatural

agents would have a privileged effect, except perhaps that escaping the monitoring of (but also the punishment from) a Big God who knows everything at any point in time is necessarily harder than from other gods, and this might lead to a stronger effect on cooperation. But beyond this putative effect, any belief in supernatural agents believed to be in the vicinity of an individual could, following this hypothesis, lead to a higher degree of cooperation.

It remains to be shown, especially concerning hypotheses of individual-level cooperation, whether and how specific beliefs in god(s) and other supernatural entities link with cooperation. A plausible, although speculative, link between cooperation and supernatural entities is that the degree to which individuals cooperate in a given society is caused by the degree to which they are “primed” in the society. The prediction would be, assuming no or a limited habituation effect, that being primed more often or with more arousing signals (such as minimally counterintuitive concepts; see Boyer, 2001) would lead to an increased level of cooperation. But to our knowledge, this line of research has not been pursued.

Cognitive Causal Mechanisms: Can We Get Down to the Specifics?

The effort to link prosociality with specific types of beliefs in the supernatural has a natural counterpart in laboratory research in social psychology, where randomized groups can be put to different treatments in experiments that search for causal explanations. Similarly, cognitive anthropologists conducting experimental fieldwork may be able to implement quasi-natural experiments to study the effects of certain variables related to religiosity, including belief, on behavior.

Although vindicating the social utility of the fear of supernatural punishment is an idea that can be traced back several centuries, its much more recent revival in cognitive science may have been spurred by a series of studies on the prosocial effects induced by minimal implicit cues of being watched. In a prototypical study on the effect of implicit social monitoring, a participant is brought to a setting in which there are subtle cues in the background that may remind one of being watched (e.g., a pair of eyes in the wallpaper of a computer screen, a photograph of two eyes posted on a board). The treatment is often considered successful if, in the subsequent measured behavior (the dependent variable of the experiment), their conduct appears more prosocial (e.g., offering money to a third party, cheating less if given the opportunity) as compared with the participants in control groups (e.g., Bateson et al., 2006; Haley & Fessler, 2005; Mifune et al., 2010; see also Bourrat, Atkinson et al., 2011, in which the proxy for prosociality is not a monetary reward but moral compliance). Extrapolating from these

results, it could be expected that cultural representations, such as external cues that served the function of being watched by some form of supernatural agent, could have a similar effect. In developmental psychology studies, the observed effect of putative supernatural “watchers” on the behavior of children was announced as an early seed of a naturally exploitable disposition to pay attention to these kinds of cues (Bering 2006; Piazza, Bering, & Ingram, 2011).

Different forms of religious priming have thus been applied as the independent variable in dozens of randomized controlled studies to elicit some form of measurable prosocial effect. A recent meta-analysis (Shariff, Willard, Andersen, & Norenzayan, 2016) found that the effect size of religious priming on prosocial behavior in these types of studies oscillated near a Hedge’s $g = 0.27$, $p < .001$, 95% CI [0.15, 0.40]. Taking possible publication bias into account resulted in $g = 0.18$, $p = .001$, 95% CI [0.04, 0.32]. It was also found that larger studies showed, on average, smaller effects.

It has been claimed that the best explanation for the observed effects in religious priming studies could be the effect of believing one is being watched by supernatural agents. Even if we accepted that other explanations are ruled out, further questions remain. If monitoring (and potentially fear) and reputational psychology are operating here, one could in principle study causation. This leads to the following questions: Are mental representations about supernatural agents in some sense causing prosociality? Or is prosociality simply an effect of being reminded that there are other fellow believers who may enforce the norm? If the former, then what are the necessary requisites for a specific belief in the supernatural to fulfill this function? And can this function be fulfilled in the absence of other contextual elements (Viciano, Loverdo, & Gomila, 2016)? In some priming experiments, the causal link between the observed increase in prosociality and the belief in a “watching” supernatural agent is not straightforward. Specifying the aspects of the religious belief that, ex-hypothesis, could be effective may be even more difficult. It is, however, crucial to understand that “religious” priming can include different types of cognitive processes. Even if we restrict ourselves to the effects of holding certain beliefs, specifying the characteristics of those beliefs can be seen as a legitimate goal if one wants to articulate a thesis linking religious belief and cooperation.

The nature of the element that does the priming and its possible dissociation with other similar cues is key to pinpointing the cognitive causes at work. Sometimes the priming can be explicit, as when making participants read a certain passage of a sacred text (Carpenter & Marshall, 2009). At other times it can be implicit, as when making participants unscramble sentences including references to the words “divine,” “sacred,” “spirit,” or “God” (Shariff & Norenzayan, 2007). Still, in all these cases the possible causal pathways to explain the observed effects can be multiple. Ara Norenzayan’s team and colleagues, for instance, have attempted to falsify the hypothesis that the observed

effects in the priming studies can be produced through simple associational “ideomotor” processes (Bargh, Chen & Burrows, 1996). Even if their arguments are compelling, there are still other possible causal hypotheses that could account for the observed effects in terms of cognitive processes. For instance, Bourrat and McKay, in an unpublished study, tried to disentangle the fear of supernatural monitoring hypothesis (in which the monitoring was not necessarily done by a divine figure) from the fear of supernatural punishment hypothesis. They detected no difference between the conditions involving priming with natural agents or priming with supernatural agents. This, in our opinion, highlights the problem of dissociating similar (but not identical) proximate mechanisms and the possibility that in many studies claiming to document a link between religion and cooperation, the real causal link does not in fact involve a belief in supernatural agency.

If the priming is based on a contextual difference anchored in the real environment where participants live, the observed effects tend to be stronger, but the possibility for disentangling the cognitive causes is even more challenging. Field studies in cognitive anthropology have sometimes found substantial effects of religiosity on measures of prosociality. For instance, Sosis and Ruffle (2003) investigated the differences to common-pool dilemmas in religious and secular kibbutzim in Israel and found that individuals in religious kibbutzim tended to claim less from the common pool, thus coordinating better in an economic game. The effect was mediated by frequency of synagogue attendance, which, in a sense, points to a Durkheimian explanation highlighting the importance of ritual for cooperation.

Similarly, Dimitris Xygalatas (2013) tested the effects of religiosity in Mauritius using common-pool dilemmas (following Sosis & Ruffle, 2003). For this study, in each pair of participants, one participant had to answer the economic dilemma in a temple (a religious setting), whereas the other participant performed the task in a restaurant (a secular setting). Those participants in the temple gave more prosocial responses by withdrawing less from the common pool, which resulted in a bigger reward for the two of them playing the economic game. They also appealed more often to “justice” while justifying their choices, although self-reported religiosity was not a predictive factor.

Erik Duhaime (2011) also studied the effect of contextual reminders of religiosity on prosocial behavior in economic games. He administered an economic dilemma (a version of the ultimatum game) to Muslim shopkeepers in Marrakech during the adhan, the very audible call to prayer in which the greatness of Allah is invoked. The study, which involved real money, revealed an important effect of this religious reminder on the prosociality of the participants. And yet, fairly allocating the real contribution of the different possible cognitive causes at work here is anything but simple. Could specific beliefs in the supernatural be acting through the mental representation of some form of a personal all-powerful agent? It is difficult to negate that these more naturalistic studies

may still point to the contribution of ritualistic participation and that only further work can settle the issue.

Lastly, how do these different psychological findings scale-up from the measures observed either in the lab or, in fewer cases, in the field, to large-scale social dynamics? (Atkinson, Latham, & Watts, 2014). Do the effect sizes observed in these studies hold at different levels of social complexity? Two different aspects would benefit from further corroboration. First, the persistence of the effect through time—as a stimuli is repeated often in time, what does this do to its effect on prosociality? There may be reasons to expect some proportional diminution of the effect, if only due to habituation (Sparks & Barclay, 2013). Second, different types of effects could reasonably be expected when the behavior in question is related to a high-stakes situation or a low-stakes situation. Similar dissociations have been found, for instance, in the literature on conformity, with normative conformity being more common, in principle, when the person is in a low-stakes situation in relation to an economic reward (Baron, Vandello, & Brunzman, 1996). In addition, the psychology of deterrence through punishment may suggest that additional factors play a role and are important: How the certainty and proximity of the supernatural punishment is cognitively calculated may play a role, since delayed, uncertain punishments normally have a weaker deterrence effect than more certain, closer in time punishments (Kleiman, 2009).

So-called Sunday effects (Norenzayan, 2013) have been reported in relation to a measured increase in cooperative tendencies (Malhotra, 2008) or a decrease in anticooperative tendencies (Edelman, 2009), as observed in highly Christian areas in the United States. This could be used in favor of the argument that, as the relevant stimuli are much stronger than in the lab and are enhanced by their actual societal context, the effects will tend to scale-up in important ways. However, the same observed Sunday effect could also be at least partly attributed to causes besides holding certain specific beliefs in supernatural agents, and it could be related to social monitoring due to engaging in the weekly ritual of attending mass on Sundays.

Before concluding, we note that the notion of cooperation is also an ambiguous one when relating it to supernatural beliefs. “Individual cooperation” and “collective-level cooperation” can be distinguished here in meaningful ways (Bourrat, Atkinson et al., 2011). The distinction is as follows: When individuals engage in collective-level cooperation, it is not straightforwardly possible to defect with respect to the benefit brought about by the cooperative behavior, whereas it is possible when individuals engage in individual cooperation. One example of collective-level cooperation is the use of money in a society. A given individual cannot reap the benefit of using money without having to be part of the cooperative game—namely, using money. Furthermore, once money is in place, any individual born in this society will be *nolens volens* using it. A

generic example of individual cooperation is when an individual does not steal someone else's goods when he has the opportunity to do so.

This distinction can be applied to the putative effects of religious belief on cooperation. In any situation in which there is no or a very limited possibility to gain a benefit from defection for individuals, there is no reason to expect natural selection to be at work at that individual level, and one will have to suppose the adaptation, if any, to be at the group or cultural level. Furthermore, cooperation will have to refer to collective-level cooperation, not individual-level cooperation. However, in some studies proposing to test individual-level fear of supernatural punishment, the proxies used to measure cooperation have been collective-level cooperation variables. For instance, Johnson (2005), testing the one byproduct–one adaptation fear of supernatural punishment hypothesis, used a variable measuring the number of jurisdictional hierarchies beyond the local community and money (two variables in the SCCS) as proxies for cooperation. One could suppose a causal link between these two variables and “individual-level” cooperation by hypothesizing that community-level cooperative “games” (such as the use of money) must have started from individual-level forms of cooperation. Thus, one could argue, even if the proxies used to measure individual-level cooperation are collective-level forms of cooperation, because collective-level forms of cooperation *initially* require individual-level forms of cooperation, they are still tracking societies in which individual-level cooperation is stronger. However, because—as we have shown—the links between religious beliefs and cooperation (both forms) are multiple and can be potentially explained in many different ways, we claim that these variables should not be regarded as good proxies for individual cooperation. Furthermore, they should be restricted, when possible, for testing hypotheses involving group-level traits.

It is true that, given the relatively low number of variables measuring cooperation in some ethnographic sources or surveys, for lack of anything better, one might be constrained to use collective-level variables of cooperation as proxies for individual cooperation. That said, when both individual cooperation variables and collective-level cooperation variables are available, if one intends to test a hypothesis at the individual level, one ought to use individual cooperation variables to avoid the aforementioned problems. This point is worth mentioning, given a recent study based on Bayesian phylogenetic methods to indirectly test two forms of the fear of supernatural punishment hypothesis in 96 Austronesian cultures (Watts et al., 2015). In this study, political complexity (measured by the number of jurisdictional hierarchies beyond the local community) was the dependent variable in the analysis, with the prediction that beliefs in moralizing gods would lead to higher levels of political complexity. They found that the presence of High Moralizing Gods does not systematically precede highly complex societies (whereas broad supernatural punishment does).

The study is particularly interesting because it may be the most serious attempt to date to control for the already mentioned Galton's problem in this area. However, the link between political complexity and individual cooperation is not immediately obvious, and without further analysis this result only yields evidence against a cultural group version of the fear of supernatural hypotheses (already an interesting result, though).¹² We hope this helps to acknowledge the multiplicity of fear of supernatural punishment hypotheses. To the credit of Watts et al. (2015), it is unclear whether the level of analysis used in this type of study (whole culture) could efficiently test any other hypothesis than a form of the cultural group level hypothesis, as Bourrat, Atkinson et al. (2011) emphasized.

Future Avenues of Research

We have shown how the study of the evolution of religious beliefs can benefit from being more explicit in terms of evolutionary mechanisms. The cultural stability of the causal cognitive chains (Sperber, 2006) leading to certain forms of religious belief can be produced by different types of processes. Not being able to pinpoint these processes can be an obstacle in the explanation of cultural phenomena. Adaptationism (both cultural and biological) can be a useful tool in the delimitation of hypotheses. Of course, corroborating an adaptationist hypothesis can be particularly challenging (as we think we have sufficiently shown). However, rejecting an adaptationist hypothesis also first requires a correct understanding of the logic behind the proposed evolutionary mechanisms, and it is not always clear what the null hypothesis should be.

Among the ground not covered in this chapter, there is room for interesting further research on the origins of religious beliefs related to the supernatural sanctioning of morality. Other alternative hypotheses we have not been able to deal with for lack of space may explain some of the interesting ethnographic associations found. First, other research in evolutionary psychology points to possible different explanations for the salience of supernatural monitoring. Psychological processes that are more domain-general than reputational psychology have been shown to produce similar effects. This has been the case, for instance, in research related to the phenomenon of unconscious vigilance and emotional arousal (Holbrook, Sousa, & Hahn-Holbrook, 2011). More directly related to the fear of supernatural punishment, even if this fear was ineffectual in terms of enforcing cooperation, error management theory (Haselton & Nettle, 2006) predicts that information that can have detrimental effects on the fitness of individuals will be, *ceteris paribus*, more salient. This can lead to this information being communicated more often, either because of its increased conversational relevance (Tofalvy & Viciano, 2009) or simply due to its increased memorability. It can also have a higher cultural persistence due to the fact that individuals give some weight to the

credibility of these beliefs. This is what a study led by anthropologist Daniel Fessler found: His team analyzed participants' acceptance of different types of statements in two online studies and found that statements framed around hazards instead of benefits tended to be more credible, a phenomenon that Fessler termed "negatively biased credulity" (Fessler, Pisor, & Navarrete, 2014). Subsequently, they analyzed two different sources of cultural data: urban legends on the Internet, and beliefs related to the supernatural in the Probability Sample Files of the Human Relations Area Files (HRAF), a collection of ethnographic reports specially selected to be representative of the wide diversity of studied human cultures. They found that beliefs having to do with the possibility of some hazard were more likely to be widespread and recorded than other beliefs. An interesting possibility is thus that the widespread diffusion of the fear of supernatural punishment in world cultures is a byproduct of this more general cognitive phenomenon (For a similar effect in cultural diffusion related to the perception of risk, see Moussaïd, Brighton, & Gaissmaier, 2015).

Other explanatory pathways not examined in this chapter deserve to be mentioned. Cultural evolutionary accounts of the prevalence of certain religious beliefs, but focusing either on different mechanisms or different religious beliefs, could be attempted. It is in principle a possibility that certain forms of religious belief are some form of evolved device for the benefit of some (but not all) individuals. In a modern form of the "religion is for the priests" outcry of the atheist Enlightenment, evolutionary mechanisms of this quasi-parasitic process could be explored. Other known cultural patterns, such as the pruning of pantheons of gods and their convergence toward a High God in the midst of political processes of ethnographic unification, as described by Robert Wright in his book *The Evolution of God* (2009), can be more difficult to formulate through general evolutionary mechanisms. However, Wright's idea that the religious belief in brotherly love tends to be developed and amplified in response to the necessities of political leaders on the ground may deserve further exploration by evolutionary-minded social scientists.

Conclusion

In this chapter, we have examined how the evolution of specific beliefs in the supernatural can be linked to the evolution of cooperation. In recent years, this area has received increasing empirical treatment. And, as we hope to have transmitted, some of the above-given findings are truly fascinating. However, we have also shown the tremendous obstacles that this enterprise has to face.

First, it is advisable to specify causal hypotheses in relation to a well-delineated evolutionary framework. Whether it is a byproduct hypothesis or an adaptationist one, it

is necessary to embed the hypothesis that one aims to test into an adequate causal narrative. Cultural adaptationism might provide a logically coherent and empirically fruitful framework, but it also offers a sober perspective on the adequacy of the current evidence in this area.

Although we have pointed to the difficulty in determining the specific content of beliefs in the supernatural in recent experimental studies, valuable estimates on the cognitive effects of activating some forms of belief in the supernatural have been produced (Sharif et al., 2015). New data analysis tools in ecological and cultural phylogenetic studies also provide us with important information to select between hypotheses (Watts et al., 2015). Nevertheless, the link between the cognitive/individual level and the group level remains somewhat elusive. Furthermore, as we have shown, these levels are sometimes not even correctly distinguished in theoretical terms, such as when group-level cooperativeness measures are presented as individual-level cooperativeness measures.

While current research tools can still be refined to better capture the elements of the specific hypotheses linking the contents of cultural beliefs and cooperation, we also believe that more intermediate level studies are missing in the current literature. The plausibility of some of the cultural phylogenetic mechanisms we discussed could increase by developing new case studies (even of the sociological or historical kind) in which the interesting ecological forces may be at play, even if these forces act over a short period of time. This future research should focus more on the situated effects of beliefs on behavior. In a sense, this could amount to some form of reconciliation between research trends focusing almost exclusively on the cognitive aspects of belief and those trends focusing almost exclusively on the more ritualistic side of religion.

References

- Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism: How to carry out an exaptationist program. *Behavioral and Brain Sciences*, 25, 489–504.
- Atkinson, Q. D., & Bourrat, P. (2011). Beliefs about God, the afterlife and morality support the role of supernatural policing in human cooperation. *Evolution and Human Behavior*, 32, 41–49.
- Atkinson, Q. D., Latham, A. J., & Watts, J. (2014). Are big gods a big deal in the emergence of big groups? *Religion, Brain and Behavior*, 5, 1–9.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*, 71(2), 230.

Baron, R. S., Vandello, J. A., & Brunsman, B. (1996). The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology, 71*, 915.

Barrett, J. L. (2004). *Why would anyone believe in God?* Lanham, MD: AltaMira Press.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*, 412-414.

Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences, 36*, 59-78.

Baumard, N., & Boyer, P. (2013). Explaining moral religions. *Trends in Cognitive Sciences, 17*, 272-280.

Baumard, N., & Chevallier, C. (2012). What goes around comes around: The evolutionary roots of the belief in immanent justice. *Journal of Cognition and Culture, 12*, 67-80.

Benedict, R. (1938) Religion. In F. Boas (Ed.) *General anthropology* (pp. 627-665). New York, Heath.

Bering, J. M. (2006). The folk psychology of souls. *Behavioral and Brain Sciences, 29*, 453-462.

Bernstein, A. E., & Katz, P. R. (2010). The rise of postmortem retribution in China and the West. *Medieval History Journal, 13*, 199-257.

Boehm, C. (2008). A biocultural evolutionary exploration of supernatural sanctioning. *The evolution of religion: Studies, theories, and critiques* (pp. 143-150). Santa Margarita, CA: Collins Foundation Press.

Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., & Gray, R. D. (2014). The ecology of religious beliefs. *Proceedings of the National Academy of Sciences, 111*, 16784-16789.

Bourrat, P. (2014). Reconceptualising evolution by natural selection (PhD dissertation). University of Sydney, Sydney.

Bourrat, P. (2015a). Levels of selection are artefacts of different fitness temporal measures. *Ratio, 28*, 40-50.

Bourrat, P. (2015b). Levels, time and fitness in evolutionary transitions in individuality. *Philosophy and Theory in Biology, 7*.

- Bourrat, P. (2015c). Origins and evolution of religion from a Darwinian point of view: Synthesis of different theories. In T. Heams, P. Huneman, G. Lecointre, & M. Silberstein (Eds.), *Handbook of evolutionary thinking in the sciences* (pp. 761–780). Dordrecht: Springer.
- Bourrat, P. (2015d). Distinguishing natural selection from other evolutionary processes in the evolution of altruism. *Biological Theory*, 7, 311–3211.
- Bourrat, P., Atkinson, Q. D., & Dunbar, R. I. (2011). Supernatural punishment and individual social compliance across cultures. *Religion, Brain and Behavior*, 1, 119–134.
- Bourrat, P., Baumard, N., & McKay, R. (2011). Surveillance cues enhance moral condemnation. *Evolutionary Psychology*, 9, 193–199.
- Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. New York: Basic Books.
- Brown, D. E. (1991). *Human universals*. New York: McGraw-Hill.
- Bull, J. J. (1994). Virulence. *Evolution*, 48, 1423–1437.
- Carpenter, T. P., & Marshall, M. A. (2009). An examination of religious priming and intrinsic religious motivation in the moral hypocrisy paradigm. *Journal for the Scientific Study of Religion*, 48, 386–393.
- Davis, W. D. (1971). Societal complexity and the nature of primitive man's conception of the supernatural (PhD thesis). University of North Carolina, Chapel Hill.
- Duhaime, E. (2011). Did religion facilitate the evolution of large-scale cooperative societies? Religious salience and the "Ritual Effect" on prosocial behavior. (MA thesis). Cambridge University, Cambridge, UK.
- Durkheim, E. (1979/1912). *Les formes élémentaires de la vie religieuse*. Paris: Presses Universitaires de France.
- Edelman, B. (2009). Red light states: Who buys online adult entertainment? *Journal of Economic Perspectives*, 23, 209–220.
- Evans-Pritchard, E. E., Firth, R., Leach, E. R., Peristiany, J. G., Layard, J., Gluckman, M., Fortes, M., & Lienhardt, G. (1954). *The institutions of primitive society: A series of broadcast talks*. Oxford, UK: Basil Blackwell.
- Fessler, D. M., Pisor, A. C., & Navarrete, C. D. (2014). Negatively-biased credulity and the cultural evolution of beliefs. *PLoS One*, 9, e95167.

- Gächter, S., & Fehr, E. (1999). Collective action as a social exchange. *Journal of Economic Behavior and Organization*, 39, 341–369.
- Gendler, T. S. (2008). Alief and belief. *Journal of Philosophy*, 105, 634–663.
- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge, UK: Cambridge University Press.
- Haley, K. J., & Fessler, D. M. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26, 245–256.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243–1248.
- Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Harvard University Press.
- Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, 10, 47–66.
- Holbrook, C., Sousa, P., & Hahn-Holbrook, J. (2011). Unconscious vigilance: Worldview defense without adaptations for terror, coalition, or uncertainty management. *Journal of Personality and Social Psychology*, 101, 451.
- Johnson, D. D. P. (2005). God's punishment and public goods. *Human Nature*, 16, 410–446.
- Johnson, D. D., & Bering, J. M. (2006). Hand of God, mind of man: Punishment and cognition in the evolution of cooperation. *Evolutionary Psychology*, 4, 219–233.
- Johnson, D., & Krüger, O. (2004). The good of wrath: Supernatural punishment and the evolution of cooperation. *Political Theology*, 5, 159–176.
- Kleiman, M. (2009). *When brute force fails: How to have less crime and less punishment*. Princeton, NJ: Princeton University Press.
- Kuran, T. (1997). *Private truths, public lies: The social consequences of preference falsification*. Cambridge, MA: Harvard University Press.
- Kurzban, R. (2011). *Why everyone (else) is a hypocrite*. Princeton, NJ: Princeton University Press.
- Malhotra, D. K. (2008, November 6). (When) are religious people nicer? Religious salience and the "Sunday effect" on pro-social behavior. Harvard Business School NOM Working Paper (09-066).

- Mifune, N., Hashimoto, H., & Yamagishi, T. (2010). Altruism toward in-group members as a reputation mechanism. *Evolution and Human Behavior, 31*, 109–117.
- Moussaïd, M., Brighton, H., & Gaissmaier, W. (2015). The amplification of risk in experimental diffusion chains. *Proceedings of the National Academy of Sciences, 112*, 5631–5636.
- Murdock, G. P. (1945). The common denominator of cultures. In R. Linton (Ed.), *The science of the man in the world crisis* (pp. 123–143). New York: Columbia University Press.
- Murdock, G. P. (1957). World ethnographic sample. *American Anthropologist, 59*, 664–687.
- Murdock, G. (1967a). *Ethnographic atlas*. Pittsburgh, PA: University of Pittsburgh Press.
- Murdock, G. (1967b). Ethnographic atlas: A summary. *Ethnology, 6*, 109–236.
- Murdock, G., & White, D. (1969). Standard cross-cultural sample. *Ethnology, 9*, 329–369.
- Naroll, R. (1965). Galton's problem: The logic of cross-cultural analysis. *Social Research, 32*, 428–451.
- Norenzayan, A. (2013). *Big gods: How religion transformed cooperation and conflict*. Princeton, NJ: Princeton University Press.
- Peregrine, P. (1996). The birth of the Gods revisited: A partial replication of guy Swanson's (1960) cross-cultural study of religion. *Cross-Cultural Research, 30*, 84–112.
- Piazza, J., Bering, J. M., & Ingram, G. (2011). "Princess Alice is watching you": Children's belief in an invisible person inhibits cheating. *Journal of Experimental Child Psychology, 109*, 311–320.
- Pinker, S., (2011). *The better angels of our nature: How violence has declined*. New York: Viking Penguin.
- Rakoczy, H., Bergfeld, D., Schwarz, I., & Fiske, E. (2015). Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development, 86*, 486–502.
- Roes, F. L., & Raymond, M. (2003). Belief in moralizing gods. *Evolution and Human Behavior, 24*, 126–135.

- Sanderson, S. K., & Roberts, W. W. (2008). The evolutionary forms of the religious life: A cross-cultural, quantitative analysis. *American Anthropologist*, *110*, 454–466.
- Schloss, J. P., & Murray, M. J. (2011). Evolutionary accounts of belief in supernatural punishment: A critical review. *Religion, Brain and Behavior*, *1*, 46–99.
- Sedikides, C., & Gebauer, J. E. (2009). Religiosity as self-enhancement: A meta-analysis of the relation between socially desirable responding and religiosity. *Personality and Social Psychology Review*, *14*, 17–36.
- Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*, 803–809.
- Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2016). Religious priming: A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review*, *20*, 27–48.
- Snarey, J. (1996). The natural environment's impact upon religious ethics: A cross-cultural study. *Journal for the Scientific Study of Religion*, *35*, 85–96.
- Sosis, R., & Kiper, J. (2014). Religion is more than belief: What evolutionary theories of religion tell us about religious commitment. In M. Bergmann & P. Kain (Eds.), *Challenges to religion and morality: Disagreements and evolution* (pp. 256–276). Oxford: Oxford University Press.
- Sosis, R., & Ruffle, B. J. (2003). Religious ritual and cooperation: Testing for a relationship on Israeli religious and secular kibbutzim. *Current Anthropology*, *44*, 713–722.
- Sparks, A., & Barclay, P. (2013). Eye images increase generosity, but not for long: The limited effect of a false cue. *Evolution and Human Behavior*, *34*, 317–322.
- Sperber, D. (2006). Why a deep understanding of cultural evolution is incompatible with shallow psychology. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 431–452). Wenner-Gren International Symposium Series. Oxford: Berg.
- Stark, R. (2001). Gods, rituals, and the moral order. *Journal for the Scientific Study of Religion*, *40*, 619–636.
- Swanson, G. E. (1964). *The birth of the gods: The origin of primitive beliefs* (Vol. 93). Ann Arbor: University of Michigan Press.

Tófalvy, T., & Viciano, H. (2009). The use of supernatural entities in moral conversations as a cultural-psychological attractor. *Annals of the New York Academy of Sciences*, 1167, 230-240.

Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. Cambridge, UK: Cambridge University Press.

Tylor, E. B. (2010). *Primitive culture: Researches into the development of mythology, philosophy, religion, art, and custom (Vol. 2)*. Cambridge, UK: Cambridge University Press. (Original work published in 1871.)

Underhill, R. (1975). Economic and political antecedents of monotheism: A cross-cultural study. *American Journal of Sociology*, 80, 841-861.

Viciano, H., & Bourrat, P. (2011). Is God an adaptation? *Philosophia*, 39, 397-408.

Viciano, H., Loverdo, C., & Gomila, T. (2016). Credibility, credulity, and redistribution: Commentary to Norenzayan et al. "The cultural evolution of prosocial religions." *Behavioral and Brain Sciences*, 39, 39-40.

Watts, J., Greenhill, S. J., Atkinson, Q. D., Currie, T. E., Bulbulia, J., & Gray, R. D. (2015). Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proceedings of the Royal Society of London B: Biological Sciences*, 282, 20142556.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, 288, 850-852.

Weeden, J., & Kurzban, R. (2013). What predicts religiosity? A multinational analysis of reproductive and cooperative morals. *Evolution and Human Behavior*, 34, 440-445.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20, 415-432.

World Value Survey (WVS). (1981-2008). World Value Survey Association: Official aggregate. .

Wright, R. (2009). *The evolution of God: The origins of our beliefs*. New York: Little Brown and Company.

Xygalatas, D. (2013). Effects of religious setting on cooperative behavior: A case study from Mauritius. *Religion, Brain and Behavior*, 3, 91-102.

Notes:

(¹) Several of the hypotheses about beliefs in supernatural entities make predictions not only for large-scale societies but also for smaller-scale ones—something that is not adequately appreciated. One of us has also published work on how the predictions may differ at the collective level and at the individual level concerning cooperation related to religious beliefs. Both types of predictions are not equally supported (see Bourrat, Atkinson et al., 2011).

(²) The reader might think at first that proposing a sharp distinction between each of these terms is largely a semantic question that has no real conceptual or empirical traction. As this chapter makes clear, distinguishing between these concepts is crucial to tease apart the mechanisms linking beliefs in supernatural entities and cooperation.

(³) Particularly relevant here is the possibility that one or several of these mental representations, or brain processes with ascribed content, can have a causal role in increasing the probability that individuals act cooperatively.

(⁴) This is a working definition for the purpose of this article. *Prima facie*, the important distinction to note here is that between *nonobservable* and *nonobserved*. Germs and other natural entities *can be* observed even if they *are not* observed. Supernatural entities here are defined as not objectively observable. To the extent that they were objectively observable, they would then be, according to this definition, natural entities. For the similarities and differences between the cognition of natural nonobserved entities such as germs, and supernatural nonobservable entities such as spirits, see Harris (2012, Ch. 8).

(⁵) Although see Weeden & Kurzban (2013) who found that the effect of religiosity on cooperative morals disappears when controlling for reproductive morals.

(⁶) Famously, Galton criticized Edward Tylor's causal linking of patrilinearity and social complexity on the basis that patrilinearity could have simply spread from earlier societies independently of the causal explanation favored by Tylor.

(⁷) In fact, under this model, when considering more than one trait, the cost c is a constraint on which it is impossible to cheat evolutionarily. This is because the benefit potentially gained by cheating would necessarily involve a cost at a later time that is greater than the benefit gained when the whole fitness of the individual is taken into account. Thus, once all costs and benefits that are inherently tied together because they originate from the same cognitive processes are taken into account, the so-called cheaters would do worse than the so-called cooperators.

⁽⁸⁾ Thus, under this hypothesis, developing a sense of fairness by natural selection imposes a constraint by subjectively increasing the cost of cheating in relation to the possible benefit earned. Not having a sense of fairness would impose overall high fitness costs due to the conditions (i.e., selective pressures) that make cooperative mutualism evolutionarily plausible.

⁽⁹⁾ Perhaps one way to separate them so would be to test whether fears of supernatural punishment are more strongly experienced in nonsocial contexts. If a difference between the social and nonsocial conditions is observed, that could be considered evidence that a cognitive mechanism specially designed to intervene in particular contexts has been selected. The rationale is that fear of supernatural punishment would be more advantageous in nonsocial contexts (or at least when individuals believe they are not in a social context) rather than in social ones in which other cognitive mechanisms already exist to manage reputation, since some experiments have shown that people are more prone to be cooperative when they know they are observed (e.g., Gächter & Fehr, 1999; Wedekind & Milinski, 2000). This is because only in contexts believed to be nonsocial by the focal individual could this individual be tempted to cheat. Besides this empirical method that would perhaps help distinguish the two hypotheses, some might consider the double-byproduct hypothesis more parsimonious than the one byproduct-one adaptation hypothesis. In fact, the rationale goes, unless one would have empirical reason not to believe so, the relation between fear of supernatural and cooperation should be regarded as resulting from a byproduct of human cognition rather than from an adaptation for reputation management.

⁽¹⁰⁾ See previous footnote for a rationale on how to distinguish both cognitive mechanisms.

⁽¹¹⁾ In this chapter, we have decided to focus on the problem of cooperation in relation to supernatural belief when it is applied to biological individuals. Thus, unless stated otherwise, we consider the costs and benefits brought about by behaviors (whether having cultural or biological origin) relative to biological individuals.

⁽¹²⁾ Roes and Raymond (2003) used a proxy for society size, with the rationale that, based on the work of Richard Alexander, believing in high moralizing gods allowed groups to become larger and decrease their probability to fission by imposing some form of impartial moral rules. But Watts et al. (2015) do not make the sort of distinction we presented in the previous section and cite Johnson (2005), who seems to test the one byproduct-one adaptation fear of supernatural punishment hypothesis. It is thus not entirely clear which evolutionary mechanism they are targeting.

Pierrick Bourrat

Pierrick Bourrat, Department of Philosophy, University of Sydney

Hugo Viciana

Hugo Viciana, University of the Balearic Islands

