

## The coevolution of sacred value and religion

Toby Handfield

To cite this article: Toby Handfield (2019): The coevolution of sacred value and religion, Religion, Brain & Behavior, DOI: [10.1080/2153599X.2019.1678512](https://doi.org/10.1080/2153599X.2019.1678512)

To link to this article: <https://doi.org/10.1080/2153599X.2019.1678512>



Published online: 27 Dec 2019.



Submit your article to this journal [↗](#)



Article views: 61




View related articles [↗](#)



View Crossmark data [↗](#)



# The coevolution of sacred value and religion

Toby Handfield 

SOPHIS, Monash University, Victoria, Australia

## ABSTRACT

Sacred value attitudes involve a distinctive profile of norm psychology: an absolutist prohibition on transgressing the value, combined with outrage at even hypothetical transgressions. This article considers three mechanisms by which such attitudes may be adaptive, and relates them to central theories regarding the evolution of religion. The first, “deterrence” mechanism functions to dissuade coercive expropriation of valuable resources. This mechanism explains the existence of sacred value attitudes prior to the development of religion and also explains analogues of sacred value in non-human species. The two remaining mechanisms, “assurance” and “coalitional,” are more likely to have been involved in the cultural evolution of specifically religious behavior. In the assurance mechanism, sacred value attitudes increase the cost of ideological commitments, making them more reliable as signals of a cooperative disposition. In the coalitional mechanism, sacred value attitudes make it dangerous for third parties to dissent from a social norm, and thus discourage competitor ideologies. While both these mechanisms are compatible with major accounts of the evolution of religion, different theories suggest a greater or lesser emphasis on one mechanism or the other.

## ARTICLE HISTORY

Received 29 April 2018  
Accepted 12 August 2019

## KEYWORDS

Sacred value; signalling;  
evolution; religion

## 1. Introduction

Two focal phenomena for an evolutionary understanding of religion are ritual and supernatural belief. Both give rise to serious puzzles. Rituals are often costly—in psychological, biological, and material terms—and lack obvious returns to fitness that would make those rituals adaptive. Given all this, why do they persist? Similarly, why are supernatural religious beliefs so prevalent? Given human cognitive systems have been subject to adaptive pressures such that they produce true beliefs about many fitness-relevant domains—folk-physics, social relations, and animal threats, for instance—why should those same cognitive systems be tolerant of beliefs that are at best underdetermined by the available evidence, and often in stark contradiction to the laws governing the visible, natural world?

While these two puzzles deserve their prominence in the field, there are other religious phenomena that warrant scrutiny. In this article, I propose an evolutionary account of a stereotypical religious attitude: the attitude of holding something to be sacred. Sacred objects, relations, or norms are imbued with an exceptionally high level of subjective value, and violations of sanctity inspire distinctive forms of condemnation and protest. Sacred valuing is a ubiquitous feature of human religious communities, and many sacred values have significant impacts on fitness. I argue there are likely to be at least three evolutionary mechanisms that have made it adaptive to hold sacred values at various points in human history, and identify some differential implications which would enable future researchers to estimate the relative significance of these different mechanisms.

The paper is organized as follows. In Section 2, I present my working definition of sacred values and a rough categorization of varieties of sacred value. In Section 3, I describe three distinct mechanisms that could explain the adaptive benefit of sacred value attitudes, and show how each mechanism gives rise to novel predictions. In Section 4, I turn to religion more broadly, and relate my account of the evolution of sacred value to what is known about the evolutionary history of religion. Section 5 concludes.

## 2. What is a sacred value?

For the purposes of this inquiry, I understand a sacred value to have two defining features:

*First*, those who uphold a sacred value will, in paradigm contexts, regard honoring that value as an absolute requirement. They will refuse to consider trading off the value against other goods, and may insist it is “beyond price.” For instance, someone who regards a particular religious object as sacred will not consent to destroy that object, even for enormous sums of money.

This apparent “pricelessness” of the sacred has to be qualified or restricted to particular contexts, however, because it is often possible to show, through choices under uncertainty, that an implicit price is ascribed, even to a sacred value. For instance, how much would a sacred valuer be willing to spend on additional insurance measures that lower the probability of their sacred object being desecrated? Presumably there is some point at which additional improvements in safety of the sacred value are no longer considered worth the additional cost, which in turn implies that there is a “price” on securing the sacred value. Despite this, it seems reasonable to focus on the central contexts—those in which explicit desecration is being considered, rather than mere remote probability of desecration—as most revealing of the distinctive attitude of a sacred valuer.<sup>1</sup>

*Second*, not only are sacred values regarded as distinctively valuable, but sacred value transgressions generate a distinctive set of moral emotions. Offering to buy somebody’s house for a very low sum may generate some irritation, but it typically won’t generate outrage. Offering to purchase something held sacred, however—be it a loss of honor, the desecration of a religious object, or the performance of a taboo behavior—is likely to generate anger, indignation, and may motivate outright retaliatory or punitive behavior.<sup>2</sup>

The outrage generated by sacred value transgressions can be further divided in terms of outrage at (i) actual or (ii) hypothetical transgressions, and whether the transgressor is (a) the sacred valuer, (b) an interlocutor, or (c) a third party (first, second, and third-person transgressions, respectively; see [Table 1](#) for a summary). In many ways the most distinctive variety of sacred outrage is experienced in response to *hypothetical first-person* transgressions: (“How dare you even suggest that I behave in that way?”) But I will cast a wide net at this stage of the inquiry, and allow that a sacred value must be one which generates any one or more of these six sub-types of outrage.

A notable omission from the above definition of the sacred is any concept of the supernatural or mystical. While there is significant interest in studying the specifically religious manifestations of sacred value attitudes, the present project requires adopting a broader definition, with a view to identifying the role of sacred value attitudes in the origins of religious behavior and organization. Incorporating the supernatural into the definition of the sacred risks trivializing the relationship between

**Table 1.** Varieties of outrage and indignation that may arise regarding sacred value transgressions.

Location of transgression	Actual	Hypothetical
First personal	“I feel so guilty/ashamed at having transgressed”	“I cannot tolerate the thought of transgressing in that way (how dare I even think of it!)”
Second personal	“I’m angry that you transgressed my sacred value”	“How dare you consider transgressing my sacred value!”
Third personal	“We should do something about the person who transgressed this value”	“I’m outraged that there are people who might transgress this value”

the two. That said, there may well be interesting further dimensions of the sacred-as-supernatural which the present inquiry will of necessity overlook.<sup>3</sup>

Having given a functional definition of sacred value attitudes in very general terms, a broad range of objects may fall under the definition. The following examples all appear to be candidate sacred values, at least in some communities. The examples are organized into three categories that will eventually be shown to have theoretical motivation: each of the three mechanisms I posit to explain the evolution of the sacred will most naturally map onto one of the three categories of sacred value object.

It will become evident from contemplating the examples discussed below that what it means to transgress a sacred value is often not derivable from knowledge of the sacred object itself. In Abrahamic cultures, cannibalism of one's deceased ancestor may be a transgression of a sacred family tie, whereas in other cultures such cannibalism could be the appropriate means to celebrate the ancestral connection. To possess a sacred value with respect to an object, therefore, is better understood as having internalized some culturally specific norms that express the relevant valuation. There may be some patterns of agreement across cultures as to what constitutes valuing/desecrating a sacred object, but this cannot be assumed in advance.

*A. Relationships to in-group members* Motivations of loyalty and devotion to friends and family are often held as sacred. Someone who puts a price on their friendship is not regarded as a true friend, and suggesting to someone that they might be willing to betray a friend, a spouse, a sibling, or a parent, will often generate intense negative, retaliatory emotions.

Not only are actual relationships imbued with sacredness, but symbols of those relationships are held sacred too. Consider suggesting to a husband or wife that they swap their wedding ring for a materially indistinguishable substitute (Medin, Schwartz, Blok, & Birnbaum, 1999); or asking a patriotic citizen to use their national flag as a dishcloth.

*B. A distribution of entitlements* Resource conflict is a ubiquitous feature of human society. When those conflicts revolve around claims to entitlements that are tied to the identity of a group, we often see examples of sacred valuation. The disputants will be uninterested in accepting any alternative compensation for forgoing the claim, and they will actively protest transgressions, both actual and hypothetical. Recent protests by indigenous Australians against tourists traversing Uluru, for instance, are not complaints that there is inadequate financial compensation being paid for this transgression: rather, the implicit claim is that Uluru should not be climbed, at any price (Hitch & Hose, 2017). Disputes over the legitimacy of Jewish settlements in the West Bank of Israel, the existence of a Palestinian right of return, and related issues are similarly liable to be examples of sacred entitlements (Ginges, Atran, Medin, & Shikaki, 2007).

In addition to these concrete entitlements which may be held as sacred, more abstract entitlements, such as equal rights or basic liberties sometimes appear to be sacred. Consider asking a committed believer in democracy whether they would tolerate an arbitrarily restricted franchise (denying women the vote, for instance) in exchange for greater economic prosperity—this sort of offer is likely to be rejected, without any attempt to consider the relative weight of the values at stake.

*C. Sacred norms* As noted above, all sacred values are characterized in terms of distinctive norms which embody the distinctive valuation attitude towards some focal object. In some cases, however, there seems to be no clear object of the valuation other than the norm itself.<sup>4</sup> Norms regarding diet, violence, sexual behavior, and various taboos tend to fall in this category—and all religions have some such norms. In contemporary western society, norms regarding sexuality tend to be among the most potent and most likely to generate outrage. Many of the most formidable insults involve allegations of deviant sexuality, and offering a stranger in a bar money to perform a taboo sexual act is probably one of the most reliable ways to trigger a violent response, at least from males.<sup>5</sup>

Apart from the absence of an intelligible value object, sacred norms can be distinguished from sacred entitlements and sacred loyalties by the *universalising* attitude that is held toward such

norms. The focus of enforcing sacred entitlements is to defend a value object from *out-group* members who may compete for the resource. The focus of enforcing sacred loyalties is to defend a cooperative relationship from free-riding *in-group* members. The focus of sacred norms, in contrast, is to uphold a norm more generally, without particular regard for group membership, but among all those who constitute the relevant moral community. This may be partly explained, at least for norms involving sex and violence, by the fact that what norm the community adopts will have significant fitness impacts on all. It also may be relevant that these norms are likely to be solving a host of coordination problems that arise among families who live together and intermarry. Once a package of solutions is found to those coordination problems, the package as a whole will have significant inertia, and cannot readily be modified in modular fashion (Sterelny, 2007, p. 325). Technologies for hunting, foraging, and violence, in contrast, can be experimented with and incrementally varied without risking disruption of the entire domestic cooperative scheme.

These three categories are not meant to be either mutually exclusive or exhaustive, but they give an initial means of organizing the space of sacred values. Some other examples appear to cross categories: a patriot might refuse to burn the flag, even for a million dollars. There seem to be grounds for thinking flag-burning is a *sacred norm* violation—it generates very broad condemnation in the relevant community. But there is also a case for thinking it a case of sacred valuation of a *relationship*—in this case loyalty to the nation. Another example is upholding the *honor* of one's family, tribe or similar group. While all human societies display some regard for honor, there are some cultures whose norms imply that honor is not merely valuable but sacred (Black-Michaud, 1975; Boehm, 1984; Miller, 1990). Honor in these cultures usually requires one or more of: hospitality, honesty, sexual chastity of females, and courage and vengefulness from males. Defense of honor motivates a number of extreme behaviors, including lethal violence. Suggesting to a member of such a culture that their honor may be foregone for money or some other non-sacred incentive would be a risky proposition.<sup>6</sup> The defense of honor may be likened to defense of an *entitlement*: losing one's standing as an honorable family will have adverse economic consequences, leading to ostracism or victimization in future. Honor is thus a sort of asset that can only be maintained by strict refusal to trade with it. But sacred valuation of honor may also be a manifestation of loyalty to one's group: often the individual who carries out a risky act of violence for defense of honor does not personally benefit from the sacrifice. So the willingness to defend honor is something done, not merely for oneself but for extended family, and the family is the primary source of shame and opprobrium for the individual who fails in this duty (Thrasher & Handfield, 2018).

### 3. Mechanisms for the evolution of sacred value

Before considering further the possible role of sacredness in the evolution of religion, consider three hypothesized mechanisms that predict adaptive benefits of sacred value attitudes. Each mechanism most readily applies to one of the particular sacred object categories described above.

#### 3.1. Deterrence (*entitlements*)

Recall that sacred valuers will not “trade” the sacred object for mere material rewards, and attempts at trade will typically generate not mere rejection, but anger. Why would it be adaptive to develop this additional emotional response?

Any trading encounter occurs in the shadow of possible defection: one's trading partner might simply steal the good they desire. This risk of robbery was even greater in the ancestral environment, where enforcement of property rights would have relied on weak institutions at best, and often be a matter of self-help.<sup>7</sup>

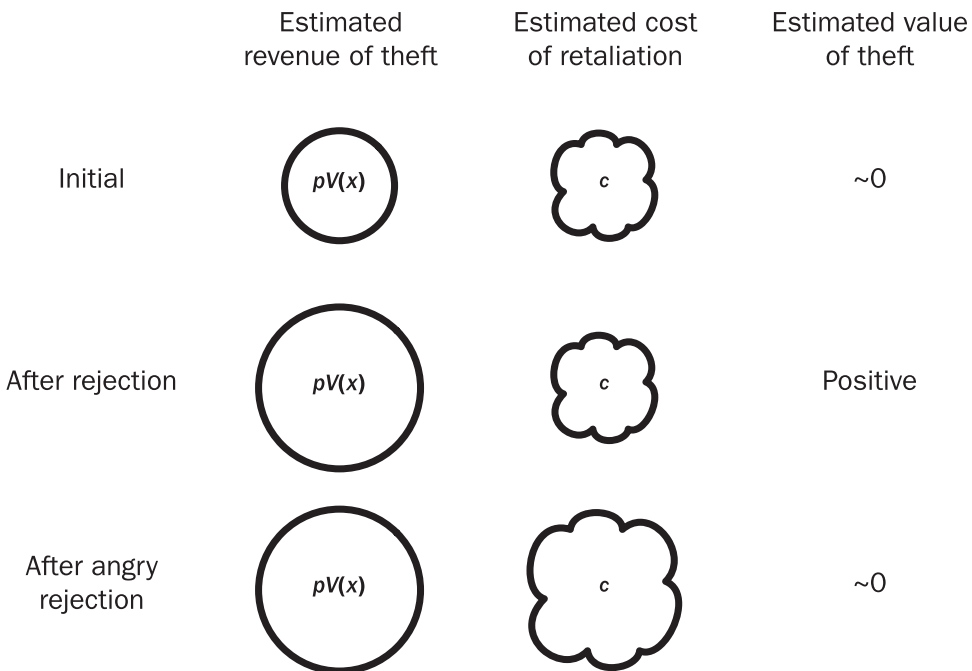
Consider a simple model of a trading partner's choice situation when contemplating whether to trade or steal. The expected value of the desired object is  $V(x)$ , which is characterized by some uncertainty—one can never know in advance the exact fitness value of an object, and the value of an object

not in one's possession is likely even more uncertain. The value of a trade can then be represented as  $V(x) - V(y)$ , where  $y$  is the item forgone in exchange for  $x$ .

The expected value of robbery, on the other hand, we represent as  $pV(x) - c$ , where  $p$  is the probability of success in obtaining the resource, and  $c$  is the expected cost of retaliation and/or conflict that will ensue.

Suppose our agent estimates that for some offered object  $y$ ,  $V(x) - V(y) > pV(x) - c$ , and also  $V(x) - V(y) > 0$ , so decides to offer the trade. If the holder of  $x$  refuses, what should the agent infer from this? First, the agent will infer that, for the other party, the values of  $x$  and  $y$  are such that the trade is a bad deal. Further, the agent may infer  $V(x)$  is greater than was first surmised. Of course, trade requires that agents have different valuations for traded goods, otherwise all exchange will be zero sum and one party will decline. But we may assume there is some positive correlation between the two parties' valuations. So the fact that the other party rejects the trade allows our agent to infer something about  $V(x)$ : our agent should update her credences such that  $x$  is judged more valuable to herself. In turn, this is likely to make robbery look like a more attractive option: the price required to purchase  $x$  is greater than initially estimated, and the benefit from possessing  $x$  is also going to be greater than initially estimated.

So refusing a trade is inherently risky: it sends a signal that might prompt more aggressive action by the trading partner. In this context, what is the effect of manifesting *outrage* at the very prospect of trade? It signals the cost  $c$  of theft is likely to be high also. It is thus a way of trying to deter possible predatory responses that may follow from a refusal to trade (see Figure 1). This hypothesis for the adaptive benefit of sacredness attitudes fits very naturally with sacred value attitudes towards *entitlements*: these are precisely the sorts of entity which give rise to both opportunities for trade and the threat of expropriation by force. Consistent with this idea, perhaps the paradigm example of a sacred entitlement is a claim to land: a resource that is often the object of fierce competitive pressure.



**Figure 1.** What an agent infers from a refusal to trade. Mere rejection of a trade makes theft more attractive, because it signals that the object in question is more valuable than initially estimated. Angry rejection sends a signal that attempted theft will be met with harsh retaliation, and hence reduces the risk of theft.

Relatedly, sacred valuers appear to perceive events central to their sacred values as closer in time—past or future (Sheikh, Ginges, & Atran, 2013). Although this phenomenon has not been demonstrated to affect subsequent decision-making, it is plausible that such temporal perceptions promote a relatively low temporal discount rate in decisions relevant to the sacred value. In turn this would imply a willingness to endure protracted conflict for the sake of ultimately securing a contested entitlement. In models of ongoing conflict, the decisive trait for victory is having a low discount rate on the future—being patient, in other words (Hammerstein & Parker, 1982). Hence the psychology of sacred values appears an excellent fit for this sort of strategic environment.

### 3.2. Coalitional (sacred norms)

Paradigm bargaining encounters involve fixed parties of determinate strength, competing for a given resource. Many social conflicts, however, depend crucially on a prior coordination issue: who will align themselves with whom? Once the composition of the relevant coalitions is evident, there will often be no need for actual conflict: the weaker side will simply capitulate.

DeScioli and Kurzban (2013) have argued that many of the distinctive features of moral codes can be explained by understanding them as a social technology to manage the problem of coalition-formation in conflict. Consider two simple strategies for coalition formation: *band-wagoning* and *alliance-seeking*. The band-wagoner, seeing a nascent dispute, chooses to side with whomever they believe would win the conflict. So in any conflict among band-wagon strategists, the stronger, bigger, more powerful individual will tend to attract allies, and the weaker party will be bound to capitulate. This sort of coalitional strategy, if widely adopted, will generate a hierarchical society, in which the pecking order determines who will recruit allies, and hence who will win any conflicts that arise. It also generates a strong temptation to despotic behavior: the strongest individual can pick a dispute with anyone, confident of the necessary support to win.<sup>8</sup> The alternative, alliance-seeking strategy is to side with whomever will most likely return the favor of support in future conflicts. This strategy requires more cognitive sophistication, because it entails weighing up future prospects and alternative strategic configurations. The outcome of using this strategy is less predictable also. DeScioli and Kurzban suggest alliance-seeking is likely to lead to a society more akin to a Hobbesian state of nature: not necessarily a state of outright combat, but an uneasy situation of shifting alliances, constant vigilance, and political intrigue. Both strategies are relatively inefficient, and stultifying to economic production. There is no incentive to generate a surplus under a despot who will expropriate it, and while the alliance-seeking solution may lead to enough security to inspire somewhat higher levels of production, it still entails weak property rights and a high probability of mis-coordination and conflict.

An alternative is to have a set of norms that coordinate responses to conflict in a more egalitarian fashion. DeScioli and Kurzban suggest the coordination can come from an informal set of *impartial* norms, especially norms that lead to clear verdicts of who is wrongdoer and who is victim. This is the posited function of “moral” rules.

From a social perspective, this solution has obvious advantages. If the rules reliably achieve coordination, they will deter victimization and theft much more effectively: any prospective transgressor will know she is likely to face the wrath of the entire community. Property rights will be more secure, and society should be more peaceful overall. That said, it may not be obvious why such an alternative should be evolutionarily viable: how does it benefit individuals to undertake the enforcement of impartial norms on third parties? Two lines of existing research help to address this puzzle without invoking cultural group selection (which may provide yet a third means of explanation).

First, agent-based models demonstrate that third parties may benefit from redistribution of resources from high fitness individuals to the less fit (Gavrilets, 2012). This result depends on the degree to which fitness depends on relative resource allocation. If the reproductive ecology approaches a “winner takes all” dynamic, then in any given encounter between individuals with

differing fitnesses, third parties will benefit if the encounter leads to a redistribution from fitter to less fit. This can make it adaptive for an individual to sacrifice fitness in order to assist a weaker individual against a stronger bully. So impartial norms that promote a more egalitarian distribution may be supported by this logic.

The second line of evidence comes from comparative ethology and anthropology. Christopher Boehm, in his study of hunter gatherer societies and their use of violence to control bullying and domineering behavior (Boehm, 1999), observes strong norms of egalitarianism. He documents explicit methods of social control and influence adopted by these communities to settle conflicts without establishing positions of political power, and from the cross-cultural uniformity of these techniques, infers they are evolutionarily stable.

DeScioli and Kurzban argue that many of the distinctive features of moral rules can be explained by the functional constraints on rules adapted for this purpose of coordinating coalition formation. For instance, the tendency of moral rules to proscribe *actions* rather than *omissions* is explained because it is much easier to detect that someone has performed a transgressive action than that they performed a transgressive omission (p. 491). They also suggest successful moral rules will tend to be absolutist, rather than allowing for casuistry and cost/benefit calculations to justify particular exceptions (p. 479). This is again because it is important that the entire community can quickly achieve consensus on who is the wrongdoer and who is the victim, without protracted investigation of the background context in which the dispute arose.<sup>9</sup> Most importantly, DeScioli and Kurzban predict this process will favor rules that are *impartial* (pp. 486–7). Adopting partisan positions, supporting those who have supported one in the past or who are likely to offer support in future, and so forth, will invite the Hobbesian variety of conflict settlement once more. Of course, kin selection will entail constant temptations to partiality, and so we should expect an ongoing tension between pressures towards partiality and pressures toward observing the society's rules in a more impartial fashion.

Recall the category of sacred norms. DeScioli and Kurzban's account lends itself to explaining several features of such norms. In particular, it explains why they are absolutist, and why they tend to prohibit particular actions that are regarded as sacrilegious, more than they tend to require positive actions. These are both predicted to arise from pressure to make the rules feasible as coordination devices in settings where information is limited and decisions must be made quickly. Further, it suggests why this category of sacred values tends to make demands more broadly than just on the immediate in-group: because these are norms that are posited to address conflicts that may arise between kin-groups, they will only be effective if upheld by a large proportion of the entire community. And the sorts of norms that are held sacred have exactly the features predicted by DeScioli and Kurzban: they tend to proscribe actions, they tend to be absolute, and they tend to be impartial. So much of the phenomenology of sacredness is explained immediately by the present account.

The outstanding question for this coalitional hypothesis is whether it explains the distinctive emotions of outrage and indignation associated with sacred values. To answer this, recall that we distinguished actual versus hypothetical transgressions and outrage directed towards transgressions that occur in the first, second, and third person. Emotional responses to second-person transgressions are straightforward to explain: if you transgress against my sacred norms, anger and outrage are presumably playing some sort of straightforward deterrence role, motivating me to undertake costly punishment for your behavior. What is more distinctive of sacred value attitudes is anger at mere hypothetical transgressions, and anger at first-person and third-person transgressions.

The prototypical first-personal emotion in response to an actual transgression against a norm is guilt or shame. A number of accounts have been offered to explain the benefit of emotions that motivate compensatory behaviors or costly apologies, as part of a process of rehabilitation and reintegration to the community (Frank, 1988; Martinez-Vaquero, Han, Pereira, & Lenaerts, 2015; O'Connor, 2016). First-personal emotions in response to mere hypothetical transgressions may be some sort of byproduct or complement of this mechanism: a means of deterring oneself from tempting transgressions.



The typical third-personal emotion, in contrast, is one of anger or outrage. Outrage in response to *actual* transgressions appears explicable in terms of the need to advertise one's willingness to undertake costly punishment, so as to coordinate the collective punishment of the transgressor. If everyone agrees who is the wrongdoer, but no one is motivated to do anything about it, the community's norms are unlikely to have much deterrent effect on future occasions. Outrage in response to *hypothetical* transgressions is much less discussed in the literature, but I suggest it is likely to serve some sort of signaling/recruiting role. In particular, where there is any risk that the community does not have a clear coordination point, by floridly disapproving of a hypothetical behavior, an individual may increase the likelihood of successful coordination in punishment of future, actual cases.

This conjecture predicts third-person hypothetical outrage (and perhaps other cases of hypothetical outrage) will be greatest in cases of norms that are weak or contested in the relevant community. In almost all societies there are very clear norms proscribing particular types of violence, for instance. Manifesting one's outrage at a hypothetical murder in such a society is not an effective signaling device, because no one is surprised at the idea of coordinating punishment against a murderer. But norms regarding homosexual behavior for instance, may be much less clearly established. Advertising one's outrage at the mere prospect of homosexuality in one's midst (or conversely, at the mere prospect of homophobia in one's midst) is thus a way to try to shape the emerging normative order in a preferred fashion. So curiously, more extreme expressions of sacred value behavior are predicted to be a feature of norms that are less well established, rather than those that have a broad consensus in the community.<sup>10</sup>

The coalitional mechanism described so far is largely a matter of sacred values deterring others from indulging competitor beliefs: "you better not believe that or else!" A second possible role for sacred valuing, however, is as a proof of sincerity. If adopting a sacred value entails personal costs because it motivates fitness-sacrificing behaviors, this constitutes powerful evidence that the professed belief is sincere. Joseph Henrich has argued the cost of religious ritual in general may be explained by a dynamic whereby:

- (1) religious ideology, when believed, makes the accompanying ritual behaviors seem less costly (because they predict divine reward);
- (2) the costliness of the ritual behaviors ("credibility-enhancing displays" or CREDs), at least to non-believers, makes it hard to think the professed belief is hypocritical or insincere.

Together, this makes the combination of belief and ritual an attractive package for a cultural learner seeking to adopt the beliefs and practices of successful agents, notwithstanding that the ritual behavior entails genuine loss of fitness (Henrich, 2009). To the extent that sacred values help translate ideological beliefs into costly CREDs, this may be a second mechanism by which sacred value contributes to stabilizing religious doctrine, by contributing to the recruitment of a coalition of supporters.

### 3.3. Assurance (relationships)

Assurance problems arise in cooperative settings where an optimal result requires individual choices that make us vulnerable to the choices of others. The collective outcome may not be achieved, not because of outright selfishness, but through a mere lack of mutual trust. The challenge then is to achieve mutual assurance of each other's reliability in performing the group beneficial behavior. Cooperation in a stag hunt is the standard example from game theory, in which by pursuing the optimal outcome (hunting stag), I make myself vulnerable to the behavior of others, but by taking the safe option (hunting hare), I can guarantee a modest payoff regardless of what others do.<sup>11</sup>

If we suppose there are a diversity of types in the population, some of whom are more motivated by the collective good than others<sup>12</sup>, then it may be possible to develop a signal that one can be counted upon to cooperate. Behaviors that are costly, public, and complements to group

participation are particularly good candidates because they will be less costly for communal individuals than for selfish individuals, and hence selfish types will be unwilling to send the signal and can thereby be “priced out” of the communal activity (Skyrms, 2010; Spence, 1973; Zahavi, 1975). An extreme example: burning the net used for catching hares is a prohibitive activity for someone who plans to hunt hare, but is no cost at all for someone who plans to hunt stag (see Figure 2). If the potential stag hunters observe the other hunters burn their nets before beginning, they can be much more confident of a successful stag hunt.

Unlike the Deterrence hypothesis and the Coalitional hypothesis, which apply to competitive interactions, this is a *cooperative* setting, where the concern is to minimize free riding, rather than to best a competitor.

Costly signaling hypotheses have been developed to explain a number of religious behaviors, but not sacred values in particular (Bulbulia, 2004; Irons, 1996; Sosis & Alcorta, 2003). In the typical case, signals are overt actions, such as the sacrifice of an animal, or participation in a painful, time-consuming, or otherwise costly ritual. These behaviors are very well suited to a signaling role because they are easy to observe, hard to fake, and highly salient. Sacred value attitudes, understood as absolute *refusals* to undertake a certain sort of behavior, are apparently a poor fit for this sort of signaling account. Failure to do a particular thing is not readily observed, more easily faked (one can try to perform the transgression later, on the sly), and is unlikely to be as salient as an overt ritualized performance.<sup>13</sup>

A related model that may be more helpful for understanding the possible signaling role of sacred value prohibitions is a *screening* account. Screening models are much like signaling models, except that where in a signaling model the signaler “moves first” by sending a signal to which they hope the receiver will favorably respond, in a screening model the receiver is first mover: setting up some obstacle that the sender must pass in order to obtain favorable treatment. Application fees (for graduate programs, social clubs, etc.) are a good example. The requirement to pay a fee will reduce the incentive for low quality individuals to apply, because they have a lower prior probability of being selected. Thus although the fee will reduce the number of applicants, it will increase the average quality of candidates to choose from.

In the religious setting, Lawrence Iannoccone predicts religious orders will develop screening mechanisms of this sort with the aim of increasing the average level of religious commitment in the group, even at cost of making the group smaller (Iannaccone, 1992). Effective screening mechanisms are requirements that *complement* religious participation and are economic *substitutes* for life

		Stag	Hare			Stag	Hare
a	Stag	4,4	0,1		b	4,4	0,1
	Hare	1,0	1,1		Stag	0,0	0,1
	Stag				Hare		

**Figure 2.** An assurance signal in a stag hunt game. In (a), the players face a standard stag hunt, in which (Stag, Stag) is a payoff dominant equilibrium, while (Hare, Hare) is risk dominant. In (b), we see the transformed payoffs after the row player has destroyed the net they would use to catch Hare. This is an activity that would be costly if row player intended to hunt Hare, but has no impact on that player’s payoff if she hunts Stag. The differential costs allow this to serve as a compelling signal that she intends to hunt Stag, increasing the likelihood that the two agents achieve the payoff dominant equilibrium.

outside the religious group—this makes them relatively “cheap” for pious individuals and “expensive” for those who are less committed. An example might be the requirement to wear stigmatizing symbols of group identity. Such requirements are much more burdensome on someone who wishes to maintain an active social life outside and to participate less fully in the religious community. Dressing in ceremonial robes, wearing an eccentric style of hair, and refusing to drink alcohol are all rather unappealing for someone who wishes to go to a nightclub, but are much less burdensome for someone who wishes to remain with the congregation to spend their leisure hours.<sup>14</sup>

This explains the role of all manner of absolute prohibitions that may be sacred value attitudes: dietary restrictions, manners of dress, requirements to undertake prayer at prescribed times. Absolutism is explained as a simple matter of monitoring costs: it is much more difficult for the community to enforce a rule that allows occasional exceptions than to enforce an absolute prohibition. But this explanation does not obviously explain the second feature of sacred value attitudes: why should they lead to distinctive moral emotions?

It may be this is an instance where, even after screening has occurred, there is still a strategic incentive to engage in additional signaling of one’s virtues.<sup>15</sup> Two acolytes may both conform with the requirements to dress in a particular way, to restrict their diets, and so forth, but one may strive to differentiate himself by additional demonstrations of zeal. By manifesting public outrage at hypothetical or actual transgressions, one is sending some sort of signal of one’s reliability and commitment. Although this signal might appear easy to fake, our disproportionately harsh judgment of hypocrites suggests that moral condemnation is a risky strategy that can therefore be a reliable signal of good character (Hok, Martin, Trail, & Shaw, *In press*; Jordan, Sommers, Bloom, & Rand, 2017).

In addition to first-person outrage at hypothetical violation of the sacred, third-person outrage at third-person violations may be a type of costly signal also. Sometimes a group will not only enforce a norm within their group, but create a norm of protesting against the non-compliance of *out-group* members, who may regard the rule as eccentric or erroneous. The Westboro Baptist Church’s campaign in recent years provides an example. They have picketed the funerals of US soldiers with slogans like “Thank God for crippled soldiers”, “God hates you”, “Fag troops”, and the like (Chen, 2015), to express the group’s belief that America’s misfortunes in war are a punishment from God for toleration of homosexuality. These behaviors stigmatize the group relative to mainstream society, and thus complement group participation while making out-group activities less attractive. This sort of costly activity has all the standard hallmarks of a signal of assurance: it imposes differential costs on those who are more or less committed to group participation; it is an overt action that is salient and hard to fake; and it is costly.<sup>16</sup>

### 3.4. Summary and differentiating predictions

The three mechanisms described above all predict that some varieties of sacred value behavior are adaptive, and each mechanism is apt to be put to further empirical test. All the mechanisms are grounded in theories that have been developed independently in a broader context, but they make distinctive predictions in the setting of sacred value. Each mechanism suggests a distinctive category of *object* that is subject to sacred value attitudes, and suggests a particular sort of social *audience* for one’s sacred value displays. These predictions are summarized in Table 2.

**Table 2.** Mechanisms by which sacred value attitudes may be adaptive, and the corresponding predicted value objects and audiences for display of these values.

Mechanism	Sacred value object	Preferred Audience
Deterrence	Entitlements	Competitors for the entitlement
Coalitional	Candidate norms, especially controversial norms	Potential norm-enforcing coalition partners
Assurance	Complements of group participation	Potential collaboration partners

#### 4. Sacred value in the evolution of religion

For these purposes, we construe religion in naturalistic, behavioral terms, as involving:

- Supernatural beliefs in agents or forces that have some concern for human affairs
- Ritualized behaviors, performed in some sort of communion. These behaviors are, at least in part, ways of expressing and communicating some of the relevant supernatural beliefs
- Costly sacrificial behaviors which are believed to have some influence over the supernatural agents

So understood, religion is a widespread, perhaps universal, feature of human communities. Dating its origin is difficult, given the absence of clear archaeological markers, but there is very little evidence to support the existence of religion prior to 100 kya, and gradually increasing evidence thereafter. In the last 10,000 years before present, we have clear evidence of organized religions in hierarchical societies that are able to construct elaborate architectural monuments such as pyramids, moai, ziggurats, and more. Sterelny (2018) argues further that religion is unlikely to have a single causal origin: it appears to be a robust feature of human communal life, readily reinvented after population shocks, unlike material technologies such as the making of fire, which have been lost to several populations in human history (Henrich, 2015, loc. 1381). So any account of the origins of religion must rely on circumstances that are relatively ubiquitous, but not overly ancient.

Sacred values appear to be particularly relevant to the development of “book” religions as opposed to “traditional” religions. This contrast is drawn in several dimensions (Hayden, 2019, pp. 5–11; Howells, 1948), but one particularly significant dimension is that of *exclusivity*. Traditional religions tend to be less hostile to other religious beliefs and traditions. Hayden goes so far as to say that there is significant innovation and experimentation in traditional religious practice (p. 15)—the opposite of what we associate with mainstream modern religions. Relatedly, traditional religions are less tightly associated with moral systems (Rappaport, 1999).

These suggest strongly that the coalitional mechanism—which functions to promote a favored norm—has been of limited relevance in the development of traditional religions, and was important primarily for book religions. A pressing question for future research is whether this can be attributed to supply side pressures (lack of social technology to enforce exclusive norms and religious uniformity) or demand side pressures (did circumstances change to make book religions a more adaptive cultural package).

Before considering the specific role of sacred values in the evolution of religion, consider two recent, prominent accounts of the evolution of religion itself. The key difference between these accounts is in the relative temporal and causal priority of the cognitive and ritualistic elements of religion.

On the “belief-first” model (Atran & Norenzayan, 2004), agents developed supernatural beliefs as a means of mastering existential anxieties which were particularly challenging in the ancestral environment. Early hominins who initiated sleeping on the ground rather than in trees were exposed to a novel range of predatory threats at night.<sup>17</sup> Environmental pressures—attributable to the substantial changes in climate that occurred towards the end of the paleolithic—are likely to have forced individuals to venture into a broader range of foraging environments, thus encountering new hazards. And increasing cognitive sophistication entailed that at some point some hominins were the first to have the ability to comprehend their own mortality.<sup>18</sup> In addition to these manifold sources of anxiety, because the presence of agentive organisms in the environment poses such a high risk, we may have evolved a specific cognitive bias to be hypersensitive to traces of agency, leading to a propensity to adopt agentive explanations even on very thin evidence (Guthrie et al., 1980). These, and possibly other factors, are hypothesized to promote supernatural accounts of natural phenomena and other proto-religious ideology.

This class of models conjectures that religious rituals developed later and some of them became especially effective at transmitting religious beliefs across generations. Someone who tells a far-

fetches a tale across the campfire may entertain, but is unlikely to inspire mass belief. But participating in a lengthy, demanding ritual as an expression of one's belief in a religious doctrine carries far more persuasive weight. By melding together belief and ritual in mutually sustaining packages, religions developed through processes of cultural selection such that only those combinations which were better adapted for faithful transmission have tended to survive.

On the "ritual-first" model, costly rituals arise without supernatural ideology, as an assurance mechanism to promote in-group cooperation. Some accounts stress the importance of change in food-gathering practices which became less mutualistic and instead relied upon temporally extended reciprocal relationships (e.g., Sterelny, 2018). Hunting bands traveled further, for longer, and upon returning, if unsuccessful, would be dependent on more successful band members to share. Foraging of this sort is high risk and cannot develop without social insurance mechanisms. Other accounts stress the challenge of intergroup conflict, which give rise to problems of free-rider management (e.g., Sosis, Kress, & Boster, 2007). But whatever the proximate challenge, all these accounts focus on ritual as a means of solving a pressing social dilemma. Rituals function as a credible signal of reliability, and thus promote cooperation.

Sterelny (2018) accounts for religious ideology in this paradigm as a side effect of the peculiarities of ritual experience. Because rituals tend to involve stressful, psychically bizarre experiences, they generate a demand for explanation. Supernatural doctrines, because they are so difficult to falsify, are then peculiarly susceptible to be adopted to explain these experiences, and become part of the package of norms and practices constituting the group's identity.

While the two types of theory disagree regarding the temporal sequence of events and the relative importance of different processes, the key causal mechanisms regarding costly rituals are compatible, and may both have been important in explaining religion's overall evolutionary trajectory. On the belief-first account, ritual behaviors signal conviction to neophytes, and thus increase the likelihood that otherwise incredible supernatural beliefs can spread to future generations with high fidelity. On the ritual-first account, the central causal role of ritual is as a technology of assurance, like a hand-shake or drinks with business colleagues.

The account to be developed below can be summarized as follows:

- (1) Deterrence type varieties of sacred value predate religion, and played little distinctive role in the evolution of religion.
- (2) To the extent that the evolution of religion was driven by forces that promoted cooperation (the ritual-first account), this suggests a significant role for sacred value to contribute via the assurance mechanism.
- (3) To the extent that the evolution of religion was driven by a process of cultural learning that did not directly address a cooperative problem (the belief-first account), the coalitional mechanism is likely to have been more important in explaining the contribution of sacred value.

#### **4.1. Entitlement-type sacred value predates religion**

My account begins with the claim that the sort of sacred value attachments associated with entitlements and a deterrence mechanism pre-date religious phenomena. There are two reasons for this assertion. First, deterrence would appear to require less cognitive sophistication than the other types of sacred value behavior. Second, while there is little or no evidence of other types of sacred valuation in other species, there is evidence of deterrence-type valuing. The most striking such evidence is the sensitivity to inequity demonstrated in a number of primate species (Brosnan & de Waal, 2003; Price & Brosnan, 2012), as well as both domesticated canines (Range, Horn, Viranyi, & Huber, 2009) and wolves (Essler, Marshall-Pescini, & Range, 2017). In the paradigm experiments, the test animal will perform a task for a modest material incentive (food, or a currency token that can be exchanged for food) but will refuse to do the same task in the presence of a conspecific that is

being paid a significantly higher “wage.” Strikingly, when refusing to work in the presence of inequity, capuchins will sometimes throw the offered wage back at the experimenter and wolves will attack the experimental apparatus. This is suggestive of “indignation,” which is partially definitive of sacred values proper.

If we grant that deterrence-type sacred value behavior was present in hominins prior to the origins of religion, this provides a precursor that may have been co-opted by the other two mechanisms. The key challenge is then to explain what pressures in the environment may have generated a demand for one or the other varieties of sacred valuing in the specific context of religion’s evolution.

#### 4.2. Religion as a solution to an assurance problem

On the ritual-first account of the evolution of religion, ritual behaviors serve as an assurance device. The selective pressure for these ritual behaviors is ascribed to a change in conditions which necessitated riskier forms of cooperation. The posited increase in demand for coordination/assurance devices fits naturally with the assurance mechanism hypothesized to explain some varieties of sacred value. Similar arguments have been made to the effect that more challenging ecologies correlate with more demanding initiation practices (Hayden, 2019, p. 104; Young, 1965) and with moralizing high gods (Botero et al., 2014).

In particular, I conjecture some elements of religious ideology associated with religious behavior became “sacralized” because this added value to the assurance function already being achieved by ritual behaviors. Note on this account, religious ritual on its own—without a sense of the sacred—may already be an assurance device. Initially, any ideology associated with that ritual, even if fervently believed, is likely to be a less effective signal, because sincere belief is hard to observe. Adopting a sacred value attitude, however, for reasons already discussed, makes belief more visible and costly.

Sacred values as means to heighten the signal value of religious participation might be particularly important in partner matching situations, where individuals are not choosing to cooperate with a large group but with a small number of select individuals. The choice of reproductive mate—one of the most fitness relevant decisions of all—is of exactly this sort. A number of links between religious behavior and sexual reproductive strategy have been documented (Slyke, 2017)—several of which suggest a role for signaling. If sexual selection is involved, it also may afford an explanation of the extremity of sacred values (Miller, 2007). If there is competition for appearance of devoutness in observation of local religious norms, it is not surprising to think that many individuals may end up placing an “absolute” value on those norms: absent other constraints, this is a compelling equilibrium solution.

By way of contrast, consider a case of assurance that occurs, apparently *without* sacred values. Sporting teams, for instance, are engaged in something like a stag hunt every time they play: showing up to the sportsfield is a little risky—if the others don’t come your time will have been wasted. Going to the cinema instead may be less rewarding, but is at least not vulnerable to the choices of others. Evidently, sporting teams very often succeed in maintaining the high risk, high reward equilibrium—so positing sacred values as a means of assurance might seem otiose.

The example is instructive in the way it fails to refute the theory. First, I suspect many sporting teams do share sacred values. Consider what would happen if the captain said she was going to leave early this week, without participating in the traditional singing of the team song after the game, because she has an opportunity to make \$200 by participating in a darts competition.<sup>19</sup> Many teams would find such a transgression outrageous—monetary compensation should not be compared to the values of solidarity with the team. Further, it is plausible that those teams who take some sort of symbols of membership to be sacred will more reliably solve the assurance problems: these teams will be able to undertake more risky ventures like extended tours, for instance. Second, sporting activities are for the most part *mutualistic* opportunities for cooperation. By running fast and hard to make a tackle, not only do I help my team-mates, I help myself too. Our interests are strongly aligned, behavior is easily monitored, and debts do not need to persist unpaid for long

periods of time. This is exactly the sort of strategic situation thought to characterize foraging and child-rearing prior to the paleolithic revolution, which could be solved by simple mechanisms of mutualism and direct reciprocity (Sterelny, 2014; Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012), so there is little wonder that an activity of this sort generates less demand for sophisticated assurance devices such as sacred values.

Ancestor cults, which become prominent in prehistoric religion during the neolithic, might be the first examples of sacred loyalties. Hayden (2019, pp. 183–5) observes that the importance of ancestor cults appears to covary with the relative wealth of the society, and—drawing on work of Maurice Freedman—suggests that ancestor worship was adaptive for family groups who possessed heritable land resources that would require collective defence from competitors. Worship of an ancestor who is associated with a particular territory appears to meet the differential cost/benefit requirements for a costly assurance mechanism. If you want to be regarded as a reliable family member, it is relatively low cost to observe rituals of respect and deference to the ancestors. If you wish to be integrated into a competitor family, it is likely to be much more costly to observe such rituals.

In sum, if religion evolved as an assurance mechanism, then the assurance mechanism account of sacred values is highly complementary. By taking religious ritual to be not just desirable but mandatory, and by making disapproval of transgression a matter of public display, the sacred valuer amplifies the signal that would be obtained by mere conformity to the ritual. The account helps to explain why religion is more than just costly rituals, but also includes ideology that is held to be sacred.

#### **4.3. Religion as the product of cognitive bias and cultural competition**

On the account that privileges cognitive processes and supernatural belief as the original basis of religion, supernatural belief is a response to management of existential dread, combined with cognitive biases that are over-eager to identify intentional agency in the environment (“hyperactive agency detection”). These elements alone predict supernatural belief, but do not impose any uniformity of belief: all manner of idiosyncratic, personal, religious doctrines may arise. Supernatural belief would be pluralistic and idiosyncratic without some sort of mechanism to coordinate and commit individuals to a common creed. This is where ritual is hypothesized to play an important role: younger generations will be more impressed by, and hence more likely to accept, beliefs that inspire costly and arduous rituals (Henrich, 2009). As already noted under the coalitional mechanism, sacred value attitudes may serve a similar function. By making belief more visible and costly, a sacred valuer will increase the apparent sincerity of those beliefs.

In addition to inspiring additional recruits, because sacred value attitudes inspire anger and indignation at the prospect of contradiction, they contribute to stabilizing the content of religious doctrine. Imbuing one’s supernatural beliefs with the attitude that contradiction warrants righteous anger is a good way of discouraging others from exploring alternative ideas. Unlike the assurance mechanism, which implies sacred value is enabling cooperation, this is a more competitive process, which may be zero- or negative sum. The belief-first account suggests this may be the more important role for sacred values.<sup>20</sup>

Again, consider a potential counterexample to this conjecture: the evolution of language. Languages evolve over time, but by and large they do not splinter into mutually incomprehensible dialects. Given each person has considerable potential for linguistic innovation, and individuals do not regard semantic and syntactic rules as matters of sacred value, how is it that languages avoid the problem that is alleged to afflict religions?

This example also fails in an instructive way. Successful linguistic communication is the solution to a coordination game, where interests are aligned. Whether we refer to this vegetable as an “eggplant” or an “aubergine”, very little is at stake for any of us: the two solutions have no differential impact on our fitness. But the sorts of outcomes that are facilitated by religious ideology very often do have fitness impacts, and conflicts of interest arise. Religious ideology is used to justify and motivate norms governing prime sites of selection pressure: marriage, reproduction, and treatment of children.

Even if the norms facilitate cooperation, it is difficult to avoid subsequent conflict over how those cooperative spoils are divided (Seabright, 2012). This entails that there is near constant pressure to renegotiate the norms and adopt alternative ones, more favorable to one party or another. Hence there is much more reason to anticipate that, without mechanisms such as sacred value attitudes as stabilizing forces, religious norms would be more prone to division and conflict.

Again, ancestor cults may illustrate an early occasion where this mechanism becomes of relevance. If there is competition within a family for control of a collective resource such as land, appealing to a loved ancestor as an authority may be a particularly effective way of bolstering one's claim to inheritance (Hayden, 2019, pp. 184–5). An extreme instance of this ideological value of ancestors is illustrated by Inka society, who regarded their emperors as immortal, and thus never actually relinquishing control of their material resources, but nonetheless determining the inheritance of power.

Because the royal mummies were not considered dead, their successors obviously could not inherit their wealth. Each Inka's *panaqa* [royal lineage] retained all of his possessions forever ... The mummies spoke through female mediums who represented the *panaqa*'s surviving courtiers or their descendants ...

After smallpox wiped out much of the political elite, each *panaqa* tried to move into the vacuum, stoking the passions of the civil war. Different mummies at different times backed different claimants to the Inka throne. After Atawallpa's victory, his *panaqa* took the mummy of Thupa Inka from its palace and burned it outside Qosqo ...

(Mann, 2006, loc. 1642–54)

To summarize this line of reasoning: if the major adaptive problem solved by religious practice is the self-reflexive one of maintaining the religion's stability and integrity through the vicissitudes of cultural evolution, then sacred values are likely to have contributed to this process by the coalitional mechanism. By advertising the norms one accepts through florid displays of outrage, and by threatening punishment for defection from those norms, sacred valuers are agents of conservation amidst potential innovation. There may well have been religions without sacred valuers in the past, but they are unlikely to have survived with their doctrines intact.

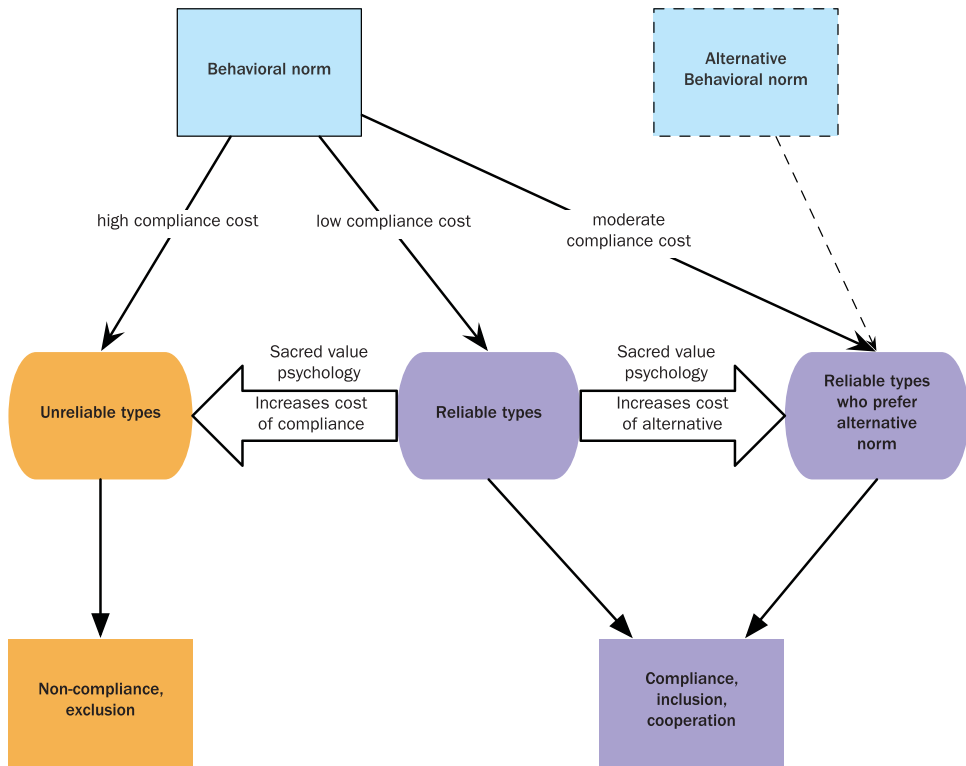
In turn, this prompts the question: what is in it for the sacred valuers themselves. Why would they care whether their local norms are preserved? Individuals will have more incentive for norm-conserving behavior when there are significant resources whose division is affected by those norms—when there is something worth contesting (recall the modeling result of Gavrillets (2012): coalitional behavior is adaptive where there is a “spoils to the victor” reproductive economy). Hence the existence of rich, defensible land resources, such as those seen in the Fertile Crescent at the beginning of the Holocene, may have been a requisite factor for this mechanism to become operative.

## Conclusion

It is simplistic to drive a sharp dichotomy between belief-first and ritual-first accounts of religion: religion is a complicated set of behaviors and it is unlikely that only a single evolutionary mechanism is involved in its development. The emerging picture is that there are two central sorts of process which might explain sacred value as instrumental in the development of religion. The assurance process is one whereby the display of sacred commitments signals to others “Trust me, I wouldn't do this if I weren't intending to stick with the group.” The coalitional process is a more competitive type of signal: “This is the norm we prefer around here, don't mess with it.” Both processes probably occurred at some point in the overall evolutionary history, regardless of temporal priority. These processes are summarized in Figure 3.

A significant shortcoming in the argument sketched here is that at various points I have appealed to the idea that sacred valuation can be an honest signal because it is a costly behavior. In some sense this is trivially so: upholding a value that requires forgoing opportunities to transgress is a “cost.” But





**Figure 3.** The roles of sacred value psychology as a coalitional device (right arrow) and as an assurance mechanism (left arrow).

this observation is not enough to show that it is costly in the right way to sustain a signaling equilibrium. Indeed, signaling models now exist in which there is arbitrarily little cost in equilibrium, but the signals are reliable because of costs that arise only in counterfactual, out of equilibrium circumstances. It is because dishonesty *would be* very costly that we can trust the costless signals we observe (Lachmann, Szamado, & Bergstrom, 2001). It is therefore pressing to develop differentially testable predictions from these theories which can be taken to the field.

Further work on this issue is not only of significance in understanding the origins of religion, but also the future of human cooperation. The role of the sacred in motivating terrorist violence behavior has been well documented (Atran, 2014), and this paper has identified a possible role for sacred values in perpetuating “culture wars” (see note 10). Given these significant costs associated with the sacred, it remains an important open question to what extent sacred value is instrumental or even essential in motivating prosocial behavior. Future empirical work will hopefully be able to distinguish the prosocial potential of sacred value mechanisms in different ecological contexts.

**Notes**

1. In terms of formal decision theory, sacred valuers probably violate (at least) the axiom of *continuity*, which states that for any three states of affairs,  $A, B, C$ , the agent strictly prefers  $A > B > C$ , if and only if there are probabilities  $p, q$ , such that the agent prefers a lottery which offers  $A$  with probability  $p$ ,  $C$  otherwise (denoted  $ApC$ ), over  $B$ , and prefers  $B$  over a lottery offering  $A$  with probability  $q$ ,  $C$  otherwise. That is,  $ApC > B > AqC$ . Consider, for example, the following states of affairs:

- $A$  = earn \$10
- $B$  = earn \$0
- $C$  = desecrate a sacred object

While it seems very plausible that a sacred valuer would rank  $A > B > C$ , there may be no probability  $p$ , such that they are willing to rank  $ApC$  over  $B$ .

Given the difficulty in providing an adequate behavioral definition of what it means to regard a value as “priceless”, it may ultimately prove more fruitful to investigate the underlying neuropsychological processes in processing sacred values. Existing evidence here is promising: it suggests that when contemplating transgressions that transgress sacred values, individuals employ distinctive regions of neural tissue, associated with deontological processing rather than weighing of costs and benefits (Berns et al., 2012).

2. Similar clusters of ideas – extremely high price and outrage at transgression – have also been referred to as “protected values” (Baron & Spranca, 1997) or “taboo tradeoffs” (Tetlock, 2003).
3. Some readers may also be interested to relate my definition of the sacred to Roy Rappaport’s influential work on ritual and sacredness in evolution (Rappaport, 1999). Rappaport reserves the word “sacred” for relatively meta-physical and unfalsifiable propositions that are foundational to a system of religious belief, and to that extent his definition is quite alien to my own (pp. 280–1). He distinguishes, however, the sacred from the “sanctified”, and allows that a very wide range of entities may be sanctified (by reference to the sacred), including norms, directives, authorities, and material objects (chap. 10). Sanctified directives or authorities – like the sacred – are “unquestionable” (chap. 9), which resonates with my understanding of sacred values as absolute. Rappaport does not appear to identify any emotions of outrage or indignation as essential to sanctification, though as I will note below, some of his speculations as to the functional role of sanctification echo elements of the mechanisms I develop.
4. Consider a borderline case, the phrase “human life is sacred”. The relevant norms that this phrase typically evokes rule out some forms of deliberate killing (murder, abortion, euthanasia), while tolerating others (violent self-defense, warfare), and having an uneasy relationship with yet others (homicide by way of negligent driving). They certainly do not enjoin a consistent and domain-general attitude of preserving human life. So while it is possible to interpret this value attitude as having an object – human life – doing so tends to obscure how very specific and circumscribed the relevant norms are.
5. In empirical research recently conducted by the author, using a culturally diverse sample of undergraduates at Monash University, participants endorsed a list of 62 value statements (derived from Berns et al., 2012), and were subsequently offered up to \$50 to sign the opposite statement. The only statement which a majority of subjects refused to sign in exchange for the available monetary incentive was one that endorsed pedophilia.
6. Honor in these cultures is, on occasion, explicitly subject to monetary transactions, though usually only with great contrivance and difficulty, and after other avenues have been exhausted. As an example, Christopher Boehm’s description of the process of negotiating a peace to end a blood feud in tribal Montenegro involves offering of significant compensation, but the compensation cannot be offered too quickly, nor accepted too quickly, lest either side incur dishonor as a result (Boehm, 1984, chap. 7). The transaction then needs to be witnessed by a disinterested council of observers, and occurs with great ceremony and banquet. Finally, after the material compensation is offered, it is typically refused by the other clan, it being both dishonorable to buy one’s way out of a blood feud, and dishonorable to accept such a transaction. Graeber (2012, p. 133) observes that *bridewealth*, the currency used for contracting marriages in some societies, is invariably distinct from the currency used for mundane transactions, symbolically denoting that ordinary money cannot be used to buy a human being. He further observes that in societies with *bridewealth*, the same special currency is used to pay blood debts.
7. Gintis, 2007; Brosnan, Jones, Gardner, Lambeth, & Schapiro, 2012 appeal to similar ideas to explain defensive attitudes towards property in chimpanzees, humans and other species.
8. Chimpanzee society is typically despotic, in this sense (Boehm, 1999).
9. This is not to say that moral rules cannot have complicated logical form – witness the bewildering complexity of attempts by philosophers to systematize intuitions about trolley cases (Thomson, 1990; Kamm, 2006). What is important is that in typical cases, there is a strong prospect of consensus as to how to categorize with respect to the rules.
10. This may go some way towards explaining the “culture wars” that are such a prominent feature of contemporary western society, despite evidence that actual divergence in political values is diminishing, at least in the US (Baker, 2005; Fiorina & Abrams, 2008). Florid outrage at transgressions of liberal values on one side, and at “political correctness” on the other, may be attempts to shape the emerging consensus on social norms of the future. New media technologies further catalyze this dynamic, making it exceptionally cheap to express outrage, and promoting the propagation of those expressions through large social networks (Crockett, 2017).
11. Repeated prisoners’ dilemmas have the same payoff structure: the low risk option is to *always* defect. The risky option is to initiate cooperation, in the hope it will be reciprocated for long enough in the future to be worth the initial sacrifice of self-interest (Skyrms, 2004).
12. Or who are less averse to the costs of contributing to the collective good – the two are equivalent at this level of analysis.
13. Though signals that are difficult to detect are in themselves a type of costly signal. A recent model demonstrates that in circumstances where the highest quality signalers are particularly choosy about who *accepts* the signal,

rather than paying to produce a more extravagant costly signal, it can be preferable to “bury” one’s signal, leading to a probability that it will never be observed (Hoffman, Hilbe, & Nowak, 2018). This provides another theoretical basis for sacred prohibitions – hard to observe sacrifices – as a costly signaling device.

14. See also Aimone, Iannaccone, Makowsky, & Rubin, 2013, which provides experimental evidence for related mechanisms to solve the problem of cooperation in a multi-player public goods game. This is important, given there has been some scepticism as to whether a signaling mechanism can solve the problem of achieving cooperation outside of 2-player signaling contexts (e.g. Henrich, 2009).
15. This is not meant to connote “virtue signaling”, which is sometimes used to derogatively describe costless signals that have little prosocial benefit.
16. Jordan, Hoffman, Bloom, & Rand, 2016 argue, in a more austere modeling context, that third-party punishment may be an effective signal of cooperativeness.
17. Skeletal evidence suggests that *Homo erectus* was similarly inept at climbing trees as modern humans, as early as 1.8 mya (Wrangham, 2009). In which case, this might suggest an implausibly early origin for religion.
18. The earliest signs of deliberate human burial date from 100 kya (Pettitt, 2013), suggesting awareness of mortality is at least this ancient.
19. Ironically, after inventing this example, I learned that a similar dispute in fact occurred between two Australian cricketers – one being the captain who wished to hasten the singing of the team song after victory to attend another social event (SBS News, 2015).
20. This is where the present account most resembles Rappaport’s ideas about the function of the sacred — or in his terminology, the “sanctified”. He notes the development of cognitive capacities for language is accompanied by enormous potential for behavioral flexibility, and that this poses a problem for social coordination. “[I]ntrinsic to increasing flexibility ... for the species ... is a problem for the separate societies ... : their members are not genetically constrained to abide by the conventions governing them, and can easily ... imagine alternatives, some of which may seem preferable to those prevailing.” He suggests “sanctity ... stabilizes the conventions of particular societies by certifying directives, authorities ... , and ... mythic discourse ... , establishing as correct particular meanings” (Rappaport, 1999, p. 321).

## Acknowledgements

Thanks to Elias Khalil, Jayani Nadarajalingam, John Morgan, William Ridge, Rich Sosis, Kim Sterelny, John Thrasher, and two anonymous reviewers for *Religion, Brain, & Behavior*, for comments on previous versions.

## Disclosure statement

No potential conflict of interest was reported by the author.

## Funding

This work was supported by Australian Research Council: [Grant Number DP150100242].

## ORCID

Toby Handfield  <http://orcid.org/0000-0003-2995-2067>

## References

- Aimone, J. A., Iannaccone, L. R., Makowsky, M. D., & Rubin, J. (2013). Endogenous group formation via unproductive costs. *The Review of Economic Studies*, 80(4), 1215–1236.
- Atran, S. (2014). *Talking to the enemy: Faith, brotherhood, and the (un)making of terrorists*. New York: HarperCollins.
- Atran, S., & Norenzayan, A. (2004). Religion’s evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and Brain Sciences*, 27(6), 713–730.
- Baker, W. E. (2005). *America’s crisis of values: Reality and perception*. Princeton, NJ: Princeton University Press.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70(1), 1–16.
- Berns, G. S., Bell, E., Capra, C. M., Prietula, M. J., Moore, S., Anderson, B., ... Atran, S. (2012). The price of your soul: Neural evidence for the non-utilitarian representation of sacred values. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1589), 754–762.
- Black-Michaud, J. (1975). *Cohesive force: Feud in the Mediterranean and the Middle East*. Oxford: Blackwell.

- Boehm, C. (1984). *Blood revenge: The anthropology of feuding in Montenegro and other tribal societies*. Lawrence, KA: University Press of Kansas.
- Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge, MA: Harvard University 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(47), 16784–16789.
- Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, 425(6955), 297–299.
- Brosnan, S. F., Jones, O. D., Gardner, M., Lambeth, S. P., & Schapiro, S. J. (2012). Evolution and the expression of biases: Situational value changes the endowment effect in chimpanzees. *Evolution and Human Behavior*, 33(4), 378–386.
- Bulbulia, J. (2004). Religious costs as adaptations that signal altruistic intention. *Evolution and Cognition*, 10(1), 19–42.
- Chen, A. (2015). Conversion via Twitter. *The New Yorker*, November 23.
- Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1(11), 769–771.
- DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, 139(2), 477–496.
- Essler, J. L., Marshall-Pescini, S., & Range, F. (2017). Domestication does not explain the presence of inequity aversion in dogs. *Current Biology*, 27(12), 1861–1865.e3.
- Fiorina, M. P., & Abrams, S. J. (2008). political polarization in the American public. *Annual Review of Political Science*, 11(1), 563–588.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: W. W. Norton & Co.
- Gavrilets, S. (2012). On the evolutionary origins of the egalitarian syndrome. *Proceedings of the National Academy of Sciences of the USA*, 109(35), 14069–14074.
- Ginges, J., Atran, S., Medin, D., & Shikaki, K. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences of the USA*, 104(18), 7357–7360.
- Gintis, H. (2007). The evolution of private property. *Journal of Economic Behavior & Organization*, 64(1), 1–16.
- Graeber, D. (2012). *Debt: The first 5000 years*. London: Penguin UK.
- Guthrie, S., Agassi, J., Andriolo, K. R., Buchdahl, D., Earhart, H. B., Greenberg, M., ... Tissot, G. (1980). A cognitive theory of religion. *Current Anthropology*, 21(2), 181–203.
- Hammerstein, P., & Parker, G. A. (1982). The asymmetric war of attrition. *Journal of Theoretical Biology*, 96(4), 647–682.
- Hayden, B. (2019). *Shamans, sorcerers, and saints: A prehistory of religion*. Washington, DC: Smithsonian Institution.
- Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior*, 30(4), 244–260.
- Henrich, J. (2015). *The secret of our success (Kindle)*. Princeton: Princeton University Press.
- Hitch, G., & Hose, N. (2017). 'It's not Disneyland': Climbing Uluru to be banned from October 2019. Retrieved from <https://www.abc.net.au/news/2017-11-01/uluru-climbs-banned-after-unanimous-board-decision/9103512>.
- Hoffman, M., Hilbe, C., & Nowak, M. A. (2018). The signal-burying game can explain why we obscure positive traits and good deeds. *Nature Human Behaviour*, 2(6), 397–404.
- Hok, H., Martin, A., Trail, Z., & Shaw, A. (In press). Children's judgements about hypocrisy: The costs and benefits of condemnation. *Child Development*.
- Howells, W. W. (1948). *The heathens: Primitive man and his religions*. New York: Doubleday.
- Iannaccone, L. R. (1992). Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy*, 100(2), 271–291.
- Irons, W. (1996). Morality, religion, and human evolution. In W. M. Richardson, & W. Wildman (Eds.), *Religion and science: History, method, and dialogue* (pp. 375–399). New York: Routledge.
- Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530(7591), 473–476.
- Jordan, J. J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science*, 28(3), 356–368.
- Kamm, F. M. (2006). *Intricate ethics: Rights, responsibilities, and permissible harm*. Oxford: Oxford University Press.
- Lachmann, M., Szamado, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences of the USA*, 98(23), 13189–13194.
- Mann, C. C. (2006). *1491: New revelations of the Americas before Columbus (Kindle)*. New York: Vintage Books.
- Martinez-Vaquero, L. A., Han, T. A., Pereira, L. M., & Lenaerts, T. (2015). Apology and forgiveness evolve to resolve failures in cooperative agreements. *Scientific Reports*, 5(10639), 1–12.
- Medin, D. L., Schwartz, H. C., Blok, S. V., & Birnbaum, L. A. (1999). The semantic side of decision making. *Psychonomic Bulletin & Review*, 6(4), 562–569.
- Miller, G. F. (2007). Sexual selection for moral virtues. *The Quarterly Review of Biology*, 82(2), 97–125.
- Miller, W. I. (1990). *Bloodtaking and peacemaking: Feud, law, and society in saga Iceland*. Chicago: University of Chicago Press.
- O'Connor, C. (2016). The evolution of guilt: A model-based approach. *Philosophy of Science*, 83, 897–908.
- Pettitt, P. (2013). *The palaeolithic origins of human burial*. London: Routledge.

- Price, S. A., & Brosnan, S. F. (2012). To each according to his need? Variability in the responses to inequity in non-human primates. *Social Justice Research*, 25(2), 140–169.
- Range, F., Horn, L., Viranyi, Z., & Huber, L. (2009). The absence of reward induces inequity aversion in dogs. *Proceedings of the National Academy of Sciences of the USA*, 106(1), 340–345.
- Rappaport, R. A. (1999). *Ritual and religion in the making of humanity*. Cambridge: Cambridge University Press.
- SBS News. (2015). *Clarke, Katich in locker room row*. Retrieved 2019, from <https://www.sbs.com.au/news/clarke-katich-in-locker-room-row>.
- Seabright, P. (2012). *The war of the sexes: How conflict and cooperation have shaped men and women from prehistory to the present*. Princeton, NJ: Princeton University Press.
- Sheikh, H., Ginges, J., & Atran, S. (2013). Sacred values in the Israeli-Palestinian conflict: Resistance to social influence, temporal discounting, and exit strategies. *Annals of the New York Academy of Sciences*, 1299(1), 11–24.
- Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge: Cambridge University Press.
- Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford: Oxford University Press.
- Slyke, J. A. V. (2017). Can sexual selection theory explain the evolution of individual and group-level religious beliefs and behaviors? *Religion, Brain & Behavior*, 7(4), 335–338.
- Sosis, R., & Alcorta, C. (2003). Signaling, solidarity, and the sacred: The evolution of religious behavior. *Evolutionary Anthropology: Issues, News, and Reviews*, 12(6), 264–274.
- Sosis, R., Kress, H. C., & Boster, J. S. (2007). Scars for war: Evaluating alternative signaling explanations for cross-cultural variance in ritual costs. *Evolution and Human Behavior*, 28, 234–247.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.
- Sterelny, K. (2007). SNAFUS: An evolutionary perspective. *Biological Theory*, 2(3), 317–328.
- Sterelny, K. (2014). A paleolithic reciprocation crisis: Symbols, signals, and norms. *Biological Theory*, 9(1), 65–77.
- Sterelny, K. (2018). Religion re-explained. *Religion, Brain & Behavior*, 8(4), 406–425.
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324.
- Thomson, J. J. (1990). *The realm of rights*. Cambridge, MA: Harvard University Press.
- Thrasher, J., & Handfield, T. (2018). Honor and violence: An account of feuds, duels, and honor killings. *Human Nature*, 29(4), 371–389.
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation. *Current Anthropology*, 53(6), 673–692.
- Wrangham, R. (2009). *Catching fire: How cooking made us human*. New York: Basic Books.
- Young, F. W. (1965). *Initiation ceremonies: A cross-cultural study of status dramatization*. Indianapolis: Bobbs-Merrill.
- Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, 53, 205–214.