Explaining moral religions

Nicolas Baumard\textsuperscript{1*} and Pascal Boyer\textsuperscript{2}

\textsuperscript{1} Institute of Cognitive and Evolutionary Anthropology, Oxford and University of Pennsylvania, Philadelphia, PA, USA
\textsuperscript{2} Washington University, St. Louis, MO, USA

Moralizing religions, unlike religions with morally indifferent gods or spirits, appeared only recently in some (not all) large-scale human societies. A crucial feature of these new religions is their emphasis on proportionality (between deeds and supernatural rewards, between sins and penance, and in the formulation of the Golden Rule, according to which one should treat others as one would like others to treat oneself). Cognitive science models that account for many properties of religion can be extended to these religions. Recent models of evolved dispositions for fairness in cooperation suggest that proportionality-based morality is highly intuitive to human beings. The cultural success of moralizing movements, secular or religious, could be explained based on proportionality.

Why religious morality?

Cognitive approaches have thoroughly renewed our understanding of religious thought and behavior [1], explaining how animacy detection systems make gods and spirits plausible [2], disgust reactions make taboos compelling [3], threat-detection systems underpin ritualized behaviors [4], coalitionary psychology strengthens membership in religious groups [5,6], memory systems constrain the organization of religious concepts [7], and intuitions about conversation inform notions of divine communication [8].

By contrast, the domain of morality has remained largely unexplored [9]. Why do some large-scale religions (although not all) come with moral prescriptions? What underlying cognitive processes explain commonalities in the moral prescriptions of otherwise very different religious systems? We propose that recent work on human cooperation and intuitive moral dispositions may provide the answer to both questions.

Moral religions are very special, and highly similar

To many Westerners, it may seem obvious that religions emphasize moral prescriptions; indeed, this may seem the very point of religion. But the evidence suggests otherwise. Most human beings for most of human evolutionary history lived in small-scale societies of foragers or horticulturalists, in which one typically finds no unified or explicit doctrine about spirits, ancestors, or gods, no established religious organization or personnel, indeed none of the other common features of religions [6]. Also, in so-called tribal religions, beliefs in spirits and gods are not usually construed as the justifications for moral prescriptions [10,11].

Recently in human evolution, and invariably in large-scale societies, there emerged organized religions, with codified ceremonies and doctrines as well as specialized personnel. In these traditions the gods were generally construed as unencumbered with moral conscience and uninterested in human morality. That is the case for the gods of classical antiquity, Sumerian, Akkadian, Egyptian, Greek, and Roman [12,13], who did not care whether people followed moral codes as long as they provided sacrifices or showed obedience. The same applies to the Aztec, Maya, and Inca gods and to classical Chinese and Hindu deities [14–17].

In some large-scale societies, at a much later stage there emerged moralizing religions. These innovative movements connected beliefs in imagined superhuman agents with highly specific moral prescriptions, such as in Jainism, Orphism, Second-Temple Judaism, and Christianity. Surprisingly, these traditions from very different places describe morality in highly similar terms. In particular, they all include prescriptions based on the same principle of proportionality.

Golden Rule

All moralizing religions defend some version of the rule that one should treat others as one would like others to treat oneself [18]. In many traditions, this principle is regarded as the core of the doctrine. Table 1 provides illustrations.

Supernatural justice

In moralizing religions, the universe is intrinsically fair. Good deeds are rewarded and misdeeds are punished, either in this world through reincarnation [17] or in the afterlife [19,20]. In either case, rewards and punishments are proportional to good and bad deeds, which is why some religious scholars, both in China [21] and in Christian Europe, imagined a special treatment or Purgatory for people who did not deserve outright salvation. Figure 1 shows illustrations of this common theme.

Penance and alms

It is assumed that misdeeds can be compensated by self-inflicted suffering (fasts, flagellation, or mutilations) or by donations to third parties (orphans, temples, or monks), the severity or generosity of which should vary with the seriousness of one’s moral transgressions [22,23].
Sainthood and heroism
Supererогatory actions, beyond the call of duty, have a special status in moral religions. In popular Buddhism for instance, saints and moral heroes accumulate good *karma* by doing more than their fair share, so they can help save sinners [17]. A similar assumption underlies Christian notions of sainthood.

Although they are generally well known to scholars, these commonalities have not been properly explained. We suggest that evolutionary dispositions for cooperation may provide the key.

Cooperation, distribution, and fairness
For a long time, the canonical situation envisaged in modeling cooperation was that of a prisoner’s dilemma, in which one cannot choose one’s partner and must select strategies that minimize the likelihood of defection [24]. However, this canonical model ignores a crucial factor in cooperation among early foragers, the fact that one can choose among potential partners, shun defectors, and favor reciprocators [25,26]. Therefore, a more appropriate model is that of a biological market [27] in which agents use signaling and reputation to convey that they are valuable cooperators [28].

Biological market models explain not just the emergence of cooperation but also the ways in which agents distribute the benefits of cooperation, a point that is crucial to human interaction but that is not explained by the classical models. Formal models show that when agents can select partners, evolutionary dynamics converge towards mutually advantageous distributions (Box 1) [29,30].

In line with this partner-choice approach, empirical evidence suggests that proportionality governs many moral intuitions.

Social exchange: proportionality between contributions and distributions
In economic games in which participants contribute to an endowment to be shared, dictators tend to adjust distributions to individual contributions, taking into account effort, talent, and investment [31–33]. Consistent with this, surveys show a widespread preference for meritocratic distributions: the more valuable someone’s input is, the more she should get [34,35].

Mutual aid: proportionality between costs and benefits
The requirement that interaction should be mutually beneficial limits the range of evolutionarily stable forms of help to those where help is mutually advantageous. In line with this prediction, experimental studies show that the more costly an action, the less stringent is the duty. Conversely, the more an action benefits the recipient, the more pressing is the obligation to accomplish it [36,37].

Punishment: proportionality between tort and compensation
Partner-choice models predict that punishment will not be aimed at deterrence (as most standard, forced-partner models imply) but rather at restoring fairness by adjusting the compensation to the tort or the punishment to the crime [38]. Experimental studies confirm that when people punish criminals, they generally downplay factors related to deterrence (likelihood of detection, publicity of the punishment, likelihood of repeat offending), intuitively taking into account the magnitude of harm and intentionality, which are relevant to restoring fairness between the perpetrator and victim [39,40]. Similarly, people readily interpret misfortune as a sanction of a misdeed, especially when the former is proportional to the latter [41]. In cooperative economic games, punishment aims at restoring fairness rather than deterring cheating [42,43]. More generally, studies show that people have strong and consistent judgments that a wrongdoer should offer compensation equivalent to the harm inflicted on the victim or, if compensation is not possible, should incur a penalty proportional to the harm done [44–46]. Obviously, proportionality (greater punishment for greater violations) all depends on people’s estimates of the respective costs of violations and sanctions. So what is an appropriate sanction to one group may seem unduly severe or lenient to another.

The preference for proportionality-based interactions appears to be universal and to develop very early. Interactions in hunter–gatherer societies require proportionality in exchange [47,48], in mutual help [47], and in punishment [38,49]. Developmental studies demonstrate that infants as young as 15 months can detect inequity [50,51] and 3-year-olds expect distributions in proportion to merit [52,53]. All these observations fit the predictions of partner-choice models.
Cultural transmission of religious morality

Evolved fairness dispositions are directly expressed in the common themes of religious morality. The Golden Rule is the simplest expression of a fairness principle, according to which (barring special conditions) one should avoid strategies that impose unnecessary costs on others. According to the principles of supernatural justice, punishment restores the balance between perpetrator and victim, leading religious scholars to imagine gradations in supernatural rewards. In the same way, fairness dispositions explain why penance and alms are intuitive to religious believers. People see moral violations as diminishing the welfare of others, so that a decrease in the welfare of perpetrators can be construed as restoring balance (even if it does not

Figure 1. Different versions of supernatural retribution. (A) Hindu judgment (19th century Indian print). The large central panel portrays Yama the God of Death seated on a throne. To the right of Yama sits Chitragupta, who keeps detailed records of every human being and on their death decides how they are to be reincarnated. (B) Chinese judgment (18th century watercolor). The Ten Kings of Hell preside over the successive spheres through which a soul must pass on its way to rebirth. (C) Christian judgment (15th century painting by Stefan Lochner). Most Christian churches teach that on Judgment Day every person who has ever lived on earth will be judged with perfect justice. (D) Buddhist judgment (19th century Tibetan tangka). The wheel of life (Tibet) comprises three inner circles, from the center outwards, depicting the poisons of ignorance, attachment, and aversion that give rise to positive and negative karma. Karma in turn gives rise to the six realms of suffering within sensible reality (samsara). The Buddha is pointing to the moon to suggest that liberation is possible. The images in this figure are in the public domain and were obtained from Wikimedia Commons.
Box 1. The evolution of fairness by partner choice

Cooperative interactions create conflicts of interest, because each partner would be better off taking a larger share of the benefit and paying a lesser cost. How can this conflict be solved? In standard reciprocity models, in which individuals cannot choose their partners, any kind of strategic advantage leads to very asymmetric offers (Figure 1A). The stronger partner indeed has a strategic advantage: the partner is forced to either accept his or her offer, or else refuse to cooperate, thereby losing all benefits.

However, human cooperative interactions occur in fluid groups [25,47] so that individuals can to some extent change partners. Those who are treated unfairly can abandon the interaction and seek a more generous partner. This creates a biological market [79] in which individuals compete to attract the best partners to be recruited in the best interactions. Agents must avoid the symmetrical pitfalls of excessive selfishness (leading partners to abandon them) and excessive generosity (allowing partners to exploit them). Simulations and analytical results demonstrate that the best evolutionarily stable strategy is to share equally the costs and benefits of the interaction (Figure 1B) [29,30]. In such models the distribution of benefits in each interaction is constrained by the whole range of outside opportuni-

ties, determined by the market of potential partners. Even in interactions in which dominant players apparently could get a larger share of the benefits, symmetric bargaining involving each partner’s outside opportunities occurs at a larger scale.

Proportionality is in most cases the best way to provide an equal share of the benefit for each cooperative partner. If, for instance, partner A has invested more than partner B, she should get a share of the benefits proportional to her investment. Otherwise partner B’s rate of return would be higher than A’s, and A would have been better off interacting with someone else. Partner choice thus explains why human interactions take on the specific form of proportionality [80].

Figure 1. (A) Evolution of the average offer by the dominant partner when individuals cannot choose their partners. (B) Evolution of the average offer by the dominant partner when individuals can choose their partners. Adapted from [30].

actually benefit the victims), but only if it is in proportion to the damage done. That is why all moral religions recommend a gradation of penance that matches differences in violations. Prescriptions for alms also engage this same logic, with the added assumption that one’s compensation for misdeeds should be proportional to one’s budget. Finally, sainthood is salient because fairness intuitions require that people do their fair share but no more, making supererogatory actions a mystery that requires an explanation.

This correspondence between universal intuitions and culturally successful representations is an instance of biased cultural transmission [54]. Cognitive accounts of human cultures take as a starting point the fact that genetic evolution produced a suite of psychological dispositions typical of modern Homo sapiens. As a result, certain kinds of information are acquired more easily and more durably than others, leading to their recurrence in many different places at different times (Figure 2) [54,55]. This explains the success of cultural productions as diverse as narratives, visual arts, the taxonomy of living things, numerical systems, and essentialist views of social groups [56]. Similarly, evolved moral intuitions provide a background against which certain types of information (in this case from moral and religious doctrines) are more easily communicated than others. Because moral intuitions often presuppose proportionality, explicit statements advocating proportionality seem more compelling and natural than possible alternatives, leading to the cultural spread of such doctrines.

Properties of religious beliefs are best explained in terms of a specific kind of dual-processing model, based on the distinction between quick, automatic, and implicit intuitive processes (often called System 1) and slow, deliberate, explicit, and general reflective processes (System 2) [57]. A variety of intuitive systems, tailored by natural selection to maximize fitness, guide behavior in a fast and automatic way [58], producing intuitive beliefs in such domains as mind-reading, artifact function, biological categories, and numerosity [59]. Human minds also entertain reflective beliefs that provide reasons for, explication of, and connections between intuitive beliefs. The main force that drives their occurrence is relevance, not truth. They are all the more stable, frequent, and culturally recurrent when they either fit our intuitions or else contradict them in an attention-grabbing manner [60]. This perspective accounts for the fact that religious beliefs are apparently diverse but thematically similar, and that they are immune to refutation and more attractive to imaginative individuals [61]. In the domain at hand, evolved moral intuitions are sufficient to nudge humans
towards cooperative behaviors, which do not require moral doctrines and explicit principles. The latter are only a form of optional a posteriori reflection on pre-existing intuitions [62].

**How is religious morality connected to human evolution?**

A cognitive description of religious morality suggests a scenario that is congruent with the experimental evidence and the historical record. In terms of cognitive processes, note that morality is not just a matter of doctrines, but of intuitions and motivations that escape conscious access. These automatic intuitions precede conscious moral reasoning, as well as the explicit justification of moral choices, which often reduces to a posteriori rationalization [63]. That is why moral intuitions are the same in religious and non-religious people [62], and adherence to particular beliefs only marginally affects prosocial behaviors [64,65]. These evolved moral intuitions, motivating individuals towards fair allocations, have supported extensive human cooperation without the need for an explanatory doctrine, religious or otherwise, in most human societies throughout evolutionary times and into the present (Box 2).

In some human groups, however, there emerged at various historical times moralizing movements in which people initiated or followed prescriptions towards a more intense adherence to intuitive moral norms (Box 3). These movements included an explicit redescription, and often a rational justification, of our common moral intuitions. Many of these movements were non-religious, because they did not link moral prescriptions to specific notions of superhuman agents. They consisted of associations based on a common interest for a secular moralistic project, as occurred in Pythagorism, Confucianism, Samkhya,

**Box 2. Did moral religions support the transition from small-scale, kin-based groups to large-scale state societies?**

Ara Norenzayan and colleagues have put forward this hypothesis and investigated its empirical basis [81,82]. The postulated development runs as follows. (i) In some groups, people imagined high gods, powerful agents that monitored people’s behavior and punished those who did not obey their rules. (ii) These high gods prescribed prosocial behaviors towards in-groups and proscribed cheating. (iii) In general, people who think they are being monitored refrain from cheating or selfish behaviors. (iv) Therefore, members of groups that had beliefs in moral gods watching them would have engaged in more prosocial behaviors, which (v) allowed these particular groups to become larger because they could avoid the problem of widespread free-riding.

Some elements of this scenario are based on solid empirical evidence. As argued for step (iii), people are indeed less likely to cheat or act in selfish ways when they assume (or are primed to think) that they are being watched [83]. However, one major problem with the scenario in our view is step (ii). As noted above, the gods of antiquity were generally not construed as being interested in people’s moral or prosocial behaviors. People did think the gods watched them, but that was to monitor the appropriate performance of rituals and sacrifices (Figure I). As a result, there seems to be no reason to assume, as in steps (iv) and (v), that believers in such gods would have been more cooperative, or that this increased cooperation would have made their societies more successful.

Indeed, the most successful ancient empires all had strikingly non-moral high gods. This is the case for Sumer, as well as for the Greek, Roman, Aztec, and Inca empires and the Mayan kingdoms. Roman history in particular would seem to suggest exactly the opposite of the proposed scenario. To simplify somewhat, the Romans, with their non-moralizing gods, built one of history’s most successful predatory empires. They then converted to Christianity, a moralizing religion, and were promptly crushed by barbarians with tribal, non-moralizing gods. Later-emerging moralizing religions could not have been the magic bullet that made large-scale societies possible, because they appeared long after the first cities and kingdoms, 5000–10 000 years too late.

**Figure I.** In most large-scale societies, people imagined amoral gods who entertained only instrumental relations with humans, either trading their help for sacrifices or punishing them for disobedience. (A) In the Sumerian mythology, Enlil helped create the humans, but then became tired of their noise and tried to annihilate them in the flood. The epic of Gilgamesh tell how a mortal known as Utnapishtim survived the flood and was made immortal by Enlil after Enil’s initial fury (Picture: Neo-Assyrian Tablet, 7th century BC). (B) Artemis, offended by the killing of a pregnant hare by Agamemnon’s soldiers, demanded that he sacrifice his own daughter Iphigenia at Aulis, as told in the play by Euripides (Picture: Mural from Pompei, 1st century BC). The images in this figure are in the public domain and were obtained from Wikimedia Commons.
Box 3. Why did moral religions emerge at the same time?

Moralizing movements and religions appeared at roughly the same time (second half of the first millennium BCE) in a few places [84,85]. Why there and then? Traditional explanations in terms of demographic change, polity size, or diffusion of agriculture are insufficient, because many large-scale societies (Egypt for instance) combined all these factors but did not give rise to moral religions.

Recent quantitative historical work may provide a better answer. Studies by Ian Morris and colleagues indicate a sharp increase in energy capture (how much energy people extract from the environment) that occurred at the same time in three distinct regions of Eurasia, the Yellow-Yangzi rivers, the Ganga valley, and the eastern part of the Mediterranean. At the end of the first millennium BCE these regions reached a production level (25 000 kcal per capita per day) that largely surpassed that of previous societies, which ranged from 4000 kcal for hunter-gatherer societies to 15 000 kcal for states such as Egypt and Uruk (Figure IA) [86,87].

These three regions are precisely the places where moral religions emerged: the Greek city states, the Gangetic Magadha and Kosala kingdoms, urban Jewish communities, and Chinese warring states (Figure IB). This suggests a tentative scenario in which the spread of moral religions followed a sharp increase in the standard of living in some Eurasian populations.

What would be the connection between these two developments? Empirical studies on the impact of economic development on individual preferences, in a variety of different cultural contexts, suggest that material prosperity allows people to detach themselves from material needs (food, protection, affiliation) [88,89]. Evolutionary reformulations of the pyramid of needs, combined with life-history theory, describe this process [90,91]. People downplay the value of higher wealth or status when these needs are met, and turn their attention to other domains of evolved preferences, such as maximizing personal wellbeing, enjoying friendship, and cultivating aesthetics, the good life that is portrayed as the goal of many moral movements [92,93]. Consistent with this, moralizing religions recruited their first adepts among the affluent social classes (e.g. the cases of Chinese Buddhism [94] and Roman Christianity [95,96]) and were all associated with asceticism or self-control techniques such as meditation that allow practitioners to detach further from materialistic drives [75,95,97–99].

![Figure 1](https://example.com/figure1.png)

**Figure 1.** (A) Evolution of energy capture since the Neolithic revolution. (B) Location of moralizing religions from 500 BCE to 200 CE: 1, Pythagoreanism; 2, Orphism; 3, Platonism; 4, Stoicism; 5, Epicureanism; 6, Christianity; 7, Second-Temple Judaism; 8, Manichaeism; 9, Ajivika; 10, Samkhya; 11, Buddhism; 12, Jainism; 13, Hinduism; 14, Confucianism; and 15, Taoism.

Platonism, and Stoicism [66,67]. In most cases, the new emphasis on explicit morality led to a change in a pre-existing religious tradition to make it more ethically oriented, as happened in Judaism [68] and Hinduism [17,69], or to the creation of new, morally focused cults, such as Jainism [70], Buddhism [17,71], Taoism [72], Orphism [17] Christianity [73,74], and Manichaeism [75].

Why did religious moral movements crowd out the secular ones, such as found in parts of Stoicism and Confucianism? Experimental evidence suggests that religious systems are more congruent with intuitions, because they describe morality in terms of the judgments of superhuman agents who have access to all relevant information about our actions, and who seem to reason in ways consistent with our own intuitive psychology [8,76]. By contrast, non-religious moral movements lack this compelling connection to intuitions. Indeed, they engage the kind of analytical, non-intuitive thinking that diminishes people’s religious commitment in experimental settings [61,77].
Despite their differences, religious and non-religious movements owe their cultural success to the fact that these explicit, coherent accounts of moral prescriptions are congruent with universal, and much older, evolved moral intuitions.

References


