

THE GOOD OF WRATH: SUPERNATURAL PUNISHMENT AND THE EVOLUTION OF COOPERATION

Dominic D.P. Johnson

Center for International
Security and Cooperation
Stanford University, Encina Hall
Stanford, CA 94305 USA
dominic@post.harvard.edu

Oliver Krüger

Department of Zoology
University of Cambridge
Downing Street, Cambridge
CB2 3EJ, UK
ok212@cam.ac.uk

ABSTRACT

Human cooperation remains a puzzle because it persists even in contexts where traditional theories predict it should not do so (i.e. among unrelated strangers, who never meet again and where reputation effects are absent). The leading explanation argues that cooperation occurs only if non-cooperators are punished. However, punishment is costly, so 'second-order' non-cooperators may arise who defect from contributing to punishment, thus unravelling this solution. We propose an alternative: during our history, the fear of *supernatural* punishment (whether real or not) deterred defectors and may therefore represent an adaptive trait favoured by natural selection. Supernatural beliefs are universal among human societies, commonly connected to taboos for public goods, so it seems plausible that they serve an important function in fostering cooperation. This hypothesis offers an explanation for (a) geographic variation in religious practices as solutions to local cooperation problems; and (b) the power of political appeals to religion to elicit cooperation.

If death were a release from everything, it would be a boon for the wicked.
Plato¹

Those who deny the existence of the Deity are not to be tolerated at all. Promises, covenants and oaths, which are the bonds of human society, can have no hold upon or sanctity for an atheist; for the taking away of God, even only in thought, dissolves all.

John Locke²

1. Plato, *The Phaedo* 107c-d.
2. John Locke, *Treatise of Civil Government and a Letter Concerning Toleration* (ed. C. Sherman; New York: Appleton-Century, 1937), pp. 212-13.

The same secret weighs heavily on our hearts...the painful secret of Gods and Kings: It's that men are free. They are free, Égisthe. You know it, and they know it not.

Jean-Paul Sartre³

Until recently, cooperation was thought to have its evolutionary origins in (a) interactions among kin,⁴ (b) reciprocated altruism between individuals that meet repeatedly,⁵ or (c) indirect reciprocity that comes from building a good reputation.⁶ Yet, none of these theories solve the puzzle of why humans continue to cooperate in large groups of genetically unrelated strangers, in single-shot interactions and when gains from reputation are negligible—that is, where all of those mechanisms are inapplicable.⁷ Consequently, a new consensus emerging from theoretical and empirical research in economics, game theory and evolutionary biology argues that cooperation will not be established unless non-cooperators (so-called ‘defectors’ or ‘cheats’) are *punished*.⁸

If punishment is so important, then the significant amount of cooperation that is evident in society must owe its success to some mechanism to detect and punish cheats. This is quite plain in today's world—we have intricate systems of laws, police to enforce these, and courts to punish defectors. However, such institutions are only very recent compared to the problem of deterring cheats over the seven million years of human history during which cooperation—at some point—evolved.⁹ Thus, it is this period of human history and pre-history on which the literature on the origins of cooperation is rightly focused: during this time, some other mechanism to deter cheats must have been necessary. But what that mechanism was remains hotly in debate.

3. The god, Jupiter, speaking to the king, Égisthe in Jean-Paul Sartre's *Les Mouches* (Paris: Gallimard, 1947).

4. W.D. Hamilton, ‘The Genetical Evolution of Social Behaviour, I & II’, *Journal of Theoretical Biology* 7 (1964), pp. 1–52.

5. R.L. Trivers, ‘The Evolution of Reciprocal Altruism’, *Quarterly Review of Biology* 46 (1971), pp. 35–57; R. Axelrod, *The Evolution of Cooperation* (London: Penguin, 1984).

6. C. Wedekind and M. Milinski, ‘Cooperation through Image Scoring in Humans’, *Science* 288 (2000), pp. 850–52; H. Gintis, E. Smith and S. Bowles, ‘Costly Signalling and Cooperation’, *Journal of Theoretical Biology* 213 (2001), pp. 103–119.

7. E. Fehr and S. Gächter, ‘Altruistic Punishment in Humans’, *Nature* 415 (2002), pp. 137–40.

8. E. Ostrom, J. Walker and R. Gardner, ‘Covenants With and Without a Sword: Self Governance Is Possible’, *American Political Science Review* 86 (1992), pp. 404–417; T.H. Clutton-Brock and G.A. Parker, ‘Punishment in Animal Societies’, *Nature* 373 (1995), pp. 209–216; E. Sober and D.S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behaviour* (Cambridge: Harvard University Press, 1998); E. Fehr and S. Gächter, ‘Cooperation and Punishment in Public Goods Experiments’, *American Economic Review* 90 (2000), pp. 980–94; Fehr and Gächter, ‘Altruistic Punishment in Humans’.

9. Sober and Wilson, *Unto Others*.

To the readers of this article, it should be clear that an important and universal aspect of human culture in both present and past times has been religion, which engenders specific, prescriptive codes of conduct. These are not 'laws', but they constitute powerful norms of behaviour. Not only are members obligated to follow these norms, religions the world over share a common crucial feature: *supernatural punishment* of those who do not. Consequently, whether real or not, supernatural punishment has a significant ability to induce cooperation. The potential effects of this deterrent have not been considered in the existing literature on the evolutionary origins of cooperation. However, recent empirical studies support the idea that it should be: Richard Sosis demonstrated that groups with costly religious beliefs signalling commitment and loyalty outlive non-religious groups, apparently due to the improved cooperation that this confers.¹⁰

Problems with Existing Theory on Punishment

Researchers agree that cooperation cannot simply rely on the *rewards* of public goods (even if large), because there still lacks a credible deterrent against defectors.¹¹ Empirical experiments bear out this claim: cooperation within real-life groups breaks down (despite higher rewards should everyone cooperate), unless cheats are punished.¹² Therefore, there is a convergence of opinion among economists, game theorists and evolutionary biologists that cooperation will emerge only if defectors are punished.¹³ However, since punishing defectors entails costs, punishment itself becomes a 'second-order public good'.¹⁴ This

10. R. Sosis, 'Religion and Intragroup Cooperation: Preliminary Results of a Comparative Analysis of Utopian Communities', *Cross-Cultural Research* 34 (2000), pp. 71-88.

11. T.C. Schelling, *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960). This reflects the fundamental paradox behind the famous paradigm of the Prisoner's Dilemma. In this game, even though each player knows that both would be better off if they were to cooperate, rational actors are expected to defect because one can never be sure that the opponent will not succumb to temptation and defect to exploit one's own cooperation. In more general analyses too, rewards turn out to be less effective *than equivalent levels* of punishment in promoting cooperation. See K. Sigmund, C. Hauert and M. Nowak, 'Reward and Punishment', *Proceedings of the National Academy of Sciences USA* 98 (2001), pp. 10757-61.

12. Ostrom, Walker and Gardner, 'Covenants With and Without a Sword'; R. Axelrod, *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration* (Princeton: Princeton University Press, 1997); Fehr and Gächter, 'Cooperation and Punishment in Public Goods Experiments'.

13. Ostrom, Walker and Gardner, 'Covenants With and Without a Sword'; Clutton-Brock and Parker, 'Punishment in Animal Societies'; Sober and Wilson, *Unto Others*; Fehr and Gächter, 'Cooperation and Punishment in Public Goods Experiments'; Fehr and Gächter, 'Altruistic Punishment in Humans'.

14. D.D. Hackathorn, 'Collective Action and the Second-order Free-rider Problem', *Rational Society* 1 (1989), pp. 78-100.

generates a new problem of 'second-order free-riders'—individuals who may indeed cooperate towards the public good, but then defect from contributing to punishment. This means that, while punishment appears to be necessary, there is no incentive for anyone to do it. So how is punishment maintained?

Three alternative solutions to this conundrum exist in the literature: (1) punishment is administered by an external institution; (2) punishment is not costly after all, or (3) both defectors and non-punishers are punished. However, none of these three mechanisms are satisfactory:

While it is useful to assume institutional enforcement in modern contexts, it leaves the evolution and maintenance of punishment unexplained because at some point in the past there were no states or institutions. Furthermore, the state plays a very small role in many contemporary small-scale societies that nonetheless exhibit a great deal of cooperative behaviour. This solution avoids the problem of punishment by relocating the costs of punishment outside the problem. The second solution, instead of relocating the costs, assumes that punishment is costless. This seems unrealistic because any attempt to inflict costs on another must be accompanied by at least some tiny cost—and any non-zero cost lands both genetic evolutionary and rational choice approaches back on the horns of the original punishment dilemma. The third solution, pushing the costs of punishment out to infinity, also seems unrealistic. Do people really punish people who fail to punish other non-punishers, and do people punish people who fail to punish people, who fail to punish non-punishers of defectors and so on, *ad infinitum*?¹⁵

Ernst Fehr and Simon Gächter recently suggested a fourth possibility, that some people become emotionally *angered* by defectors and, as a result, are prepared to endure a cost to punish them.¹⁶ However, in Fehr and Gächter's experiments all subjects were anonymous. It is easy to press a punishment button without having to identify oneself to the victim. But in real life (and certainly more so in our past), defectors would have to be confronted. A significant problem remains, therefore, because vigilantism is costly, dangerous and haphazard—people may resent the fellow citizens that punish them, introducing the risk of spite and revenge that could undermine continued cooperation.

15. R. Boyd and P.J. Richardson, 'Why People Punish Defectors: Weak Conformist Transmission Can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas', *Journal of Theoretical Biology* 208 (2001), pp. 79–89 (80).

16. Fehr and Gächter, 'Altruistic Punishment in Humans'. For the *group*, that cost paid off in subsequent games because, with punishment, cooperation became established and the mutual rewards then exceeded those from non-cooperation. However, the punishment was 'altruistic' because interactions were randomized between unknown partners and then rotated, and the cost of punishing others within any one game reduced net gains below the payoff for doing nothing. Thus, people were effectively punishing for the good of others in the community by coercing the never-to-be-seen-again defectors to invest more in such games in the future.

So whether people within a community really do impose a credible threat on *each other* as potential punishers remains suspect (empirically, we know that vigilantism is not common and is discouraged).¹⁷ The puzzle therefore remains: without institutions of law and order, and without a good incentive for people to punish each other, how could early human societies establish cooperative norms which entailed a credible deterrent threat?

An Alternative: Supernatural Punishment

We suggest that a solution to this problem during human evolution may derive from that universal characteristic of all societies: religion. Religion offered a system of cooperation enforced by punishment, without modern institutions or second-order free-riders, via three complementary mechanisms:

1. Religious traditions, taboos and mythology provided the 'laws' (the rights and wrongs which defined the norms of conduct, promoting, among other things, cooperation).
2. Religion provided the threat of supernatural punishment to enforce those norms (both in the present and/or in the afterlife). If supernatural punishment is held as a *belief*, then this threat becomes a deterrent *in reality*, meaning that the mechanism can work *regardless* of whether the threat is genuine or not. This follows Thomas's dictum: 'If men define situations as real, they are real in their consequences'.¹⁸
3. Enforcement is also supplemented by public admonishment of people who disregard the norms. In particular, ostracism from one's community was a very severe threat (and involved worldly as well as other-worldly punishment).

Religion of course has positive aspects as well (for example, reward, and redemption, for good behaviour), but the focus in the cooperation literature has centred on the significance of punishment. 'Carrots' are not enough because, although they may encourage some people to cooperate, they do not prevent all of them from cheating. Thus, rewards may contribute to promoting cooperation, but it is the weaker of the two complementary forces. Indeed, 'the very proclamation of hell indicates that the defenders of religion found it necessary to balance the attraction of its promise with a threat for the "others", who rejected it or failed to meet its tests'.¹⁹

17. D.D.P. Johnson, P. Stopka and S. Knights, 'The Puzzle of Human Cooperation', *Nature* 421 (2003), pp. 911-12.

18. W.I. Thomas and D.S. Thomas, *The Child in America: Behaviour Problems and Programs* (New York: Alfred A. Knopf, 1928), pp. 571-72.

19. A.E. Bernstein, *The Formation of Hell: Death and Retribution in the Ancient and Early Christian Worlds* (Cornell: Cornell University Press, 1993), p. x (preface).

What do we mean by supernatural punishment? A Christian, for example, who acts contrary to God's will, fears punishment now and in an afterlife. He faces divine retribution for violating the established norms of behaviour within his religious community: 'It is plain from the Bible that sin will be punished (Dan. 12:2; Matt. 10:15; John 5:28ff.; Rom. 5:12ff., etc.).'²⁰ In the Old Testament, 'sin necessarily and inevitably involves punishment', while in the New Testament, '... "punishment" is not as common as "condemnation", [but] to be condemned is sufficient. Punishment is implied'.²¹ In other words, believers who break the rules expect to suffer retribution. The concept of the afterlife means that, in addition to any immediate sanctions that may be feared (such as being struck down with an affliction or some other misfortune), those who cheat the moral codes of the community while on earth bear the threat of punishment later, in hell.

In Buddhism and the Hindu traditions, the state in which one is reincarnated is in part dependent on one's ethical behaviour in the present life. There is thus a strong incentive to behave according to their religious codes. Muslims, as another example, follow codes of behaviour in a community that also carries the moral obligation to maintain 'that which is right and to forbid that which is wrong'.²² As a final example, people who break religious codes among the Zuni of New Mexico are physically punished, and learn that they may fail to find their way to the afterlife.²³ The deterrent effect of supernatural punishment should not be underestimated, especially in the era of pre-industrial human societies when many natural phenomena remained inexplicable and belief systems, myths and folklore were often universally held truths. Archaeological excavations reveal that as much as 50,000 years ago, Cro-Magnon people conducted ceremonial burials, burying the dead with valuable jewellery, ornaments, tools and flowers. This practice is thought to represent the origins of a belief in an afterlife.²⁴ Thus, although we may never know for sure, the belief that what one does in this life has a supernatural consequence in the next, may be very ancient indeed.

Our theory raises three immediate questions, which we address below:

20. E. Harrison, G. Bromiley and C. Henry (eds.), *Wycliffe Dictionary of Theology* (Peabody, MA: Hendrickson, 1960), p. 196. First sentence under the heading 'Eternal Punishment'.

21. *Wycliffe Dictionary of Theology*, p. 430.

22. Extract is from the Koran (3:104), cited in S.H. Hashmi, 'Is there an Islamic Ethic of Humanitarian Intervention?', *Ethics and International Affairs* 7 (1993), pp. 55-73.

23. H.B. Earhart (ed), *Religious Traditions of the World* (New York: HarperCollins, 1993).

24. R. Leakey and R. Lewin, *Origins Reconsidered: In Search of What Makes Us Human* (London: Anchor, 1993).

1. We have proposed an ultimate function for supernatural punishment, but how does this fit with the proximate reasons that people cite for their religious convictions?
2. Cooperation is found in all human societies, but religions are very diverse, so is supernatural punishment a widespread feature?
3. Religion is sometimes exploited as a tool of power by the elite, so does it really provide cooperative advantages for the average individual?

Supernatural Punishment May Serve Ultimate Functions Via Proximate Mechanisms
 Alternative 'explanations' for religion, such as that it serves to seek redemption or personal salvation, are not exclusive of our hypothesis. Rather, we would concur that these emotions may represent the *proximate* stimulant that is a prerequisite of any *ultimate* adaptive behaviour. Max Weber neatly summarizes the power of supernatural punishment for ultimate (rational) outcomes, with the salvation notion as the proximate stimulus that may eventually have become desired for its own sake:

...whoever flouted divinely appointed norms would be overtaken by the ethical displeasure of the god who had these norms under his special care...transgression against the will of god is an ethical sin which burdens the conscience, quite apart from its direct results. Evils befalling the individual are divinely appointed inflictions and the consequences of sin, from which the individual hopes to be freed by 'piety' (behaviour acceptable to god) which will bring the individual salvation. In the Old Testament, the idea of 'salvation,' pregnant with consequences, still has the elementary rational meaning of liberation from concrete ills.

In its early stages, the religious ethic consistently shares another characteristic with magic worship in that it is frequently composed of a complex of heterogeneous prescriptions and prohibitions derived from the most diverse motives and occasions. Within this complex there is, from our modern point of view, little differentiation between important and unimportant requirements; any infraction of the ethic constitutes sin. Later, a systematisation of these ethical concepts may ensue, which leads from the rational wish to insure personal external pleasures for oneself by performing acts pleasing to the god, to a view of sin as the unified power of the anti-divine (diabolical) into whose grasp man may fall. Goodness is then envisaged as an integral capacity for an attitude of holiness, and for consistent behaviour derived from such an attitude. During this process of transformation, there also develops a hope for salvation as an irrational yearning to be able to be good for its own sake, in order to gain the beneficent awareness of such virtuousness.²⁵

25. M. Weber, *The Sociology of Religion* (Berkeley: University of California Press, 1978), see section 8 'From Magical Ethics to Conscience, Sin and Salvation' in the chapter 'Magic and Religion'.

Hence, the process can be self-reinforcing. Indeed, the stronger one believes, the greater the perceived costs of defection may be. St Augustine cautioned that, 'the more enjoyment man found in God, the greater was his wickedness in abandoning him; and he who destroyed in himself a good which might have been eternal, became worthy of eternal evil'.²⁶

Cooperative Norms and Supernatural Punishment are Ubiquitous

Our hypothesis predicts that supernatural punishment, if it is such a crucial method of promoting cooperation, should be present in most human societies. Religious beliefs in general are a universal feature of all human societies known from both past and present times. George Murdock first catalogued the common features that existed in all of the several hundred societies studied at that time.²⁷ Among the 67 he found were cooperative labour, religious rituals, and propitiation of supernatural beings. Punishment for flouting norms of behaviour, specifically, is indeed also a common aspect of religion across the world's cultures.²⁸ Sentiments of hostility towards defectors who behave 'unfairly' is another systematic feature of all western and non-western societies studied, from those deep in the Amazon Basin to those in New York city.²⁹ Of course, cultural norms and traditions are extremely diverse, but:

the evident variety of moral beliefs and practices among various cultures has proven to be somewhat misleading, because it masks much that cultures have in common. Comparative ethicists have shown that there are no pre-moral societies; that all societies give some degree of moral value to such things as human life, sexual restraint, friendship, mutual aid, fairness, truthfulness, and generosity; and that all societies employ moral concepts such as good, bad, right, wrong, just, and unjust.³⁰

The Power and the People

Weber also noted in *The Sociology of Religion* that religious beliefs are a constant among human groups, and that religious norms are observed by rulers as well as subjects who were, like the 'men' they ruled, subjugated to a god, even if they were in a position to extend it as a tool of exploitation:

26. St. Augustin, *City of God* 21:12.

27. G.P. Murdock, 'The Common Denominator of Cultures', in *idem* (ed.), *The Science of Man in the World Crisis* (New York: Columbia University Press, 1945).

28. Earhart (ed.), *Religious Traditions of the World*.

29. J.H. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics* (Princeton: Princeton University Press, 1995); J. Henrich, 'Does Culture Matter in Economic Behaviour? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon', *American Economic Review* 90 (September 2000), pp. 973-79; J. Henrich *et al.*, 'In Search of Homo economicus: Behavioural Experiments in 15 Small-scale Societies', *American Economic Review* 91.2 (2001), pp. 73-78.

30. D.A. Welch, *Justice and the Genesis of War* (Cambridge: Cambridge University Press, 1993), p. 195.

There is no concerted action, as there is no individual action, without its special god. Indeed, if an association is to be permanently guaranteed, it must have such a god. Whenever an organisation is not the personal power base of an individual ruler, but genuinely an association of men, it has need of a god of its own... It is a universal phenomenon that the formation of a political association entails subordination to its corresponding god.³¹

It is a common theme that certain community members assume the responsibility of monitoring and punishing defectors and, although people in such positions of power are sometimes corrupt, they are not immune from precisely the same deterrent. The following appears under 'Punishment' in the *Wycliffe Dictionary of Theology*:

Throughout the Bible it is insisted that sin is to be punished. In an ultimate sense God will see that this is done, but temporarily the obligation is laid upon those in authority to see that wrongdoers are punished. The *lex talionis* of Ex. 21:23-25 is not the expression of a vindictive spirit. Rather it assures an even justice (the rich and the poor are treated alike), and a penalty proportionate to the crime.³²

In many pre-industrial societies' religions too, it is shamans or spiritual headmen (and/or spirits or gods) who detect and punish those who flout religious norms.³³

If the average individual incurs some cost from following religious norms (even if exacerbated by an elite—or genuine gods—'imposing' religion upon them) the benefits gained from cooperative compliance may still prevail.³⁴ The potential benefits of cooperation within such bound groups are enormous because cooperative associations reap benefits far greater than the sum of indi-

31. Weber, *The Sociology of Religion*, see section 4 'Pantheon and Functional Gods' in the chapter 'The Origins of Religion'.

32. *Wycliffe Dictionary of Theology*, p. 430.

33. Earhart, *Religious Traditions of the World*.

34. In India, 'any effort to emerge from one's caste, and especially to intrude into the sphere of activities appropriate to other and higher castes, was expected to result in evil magic and entailed the likelihood of unfavourable incarnation hereafter. This explains why, according to numerous observations on affairs in India, it is precisely the lowest classes, who would naturally be most desirous of improving their status in subsequent incarnations, that cling most steadfastly to their caste obligations, never thinking of toppling the caste system through social revolutions or reform. Among the Hindus, the Biblical emphasis echoed in Luther's injunction, "Remain steadfast in your vocation", was elevated into a cardinal religious obligation and was fortified by powerful religious sanctions'. Weber, *The Sociology of Religion*, see section 7 'Caste Taboo, Vocational Caste Ethics, and Capitalism' in the chapter 'Magic and Religion'.

vidual contributions.³⁵ In early human societies, numerous such cooperative activities were crucial to survival (for example, hunting, food sharing, construction, division of labour, collective defence).³⁶ Nevertheless, a cooperative society is always open to invasion by those who cheat the system;³⁷ even the great benefits of public goods that all can see have an Achilles heel which can destroy cooperation. As written in *The Federalist Papers*, 'If men were angels, no government would be necessary.'³⁸ Unfortunately, men (or women) are not angels. There is always the temptation to free-ride off the benevolence of other cooperators without contributing anything back. Numerous experiments, computer simulations and real-life case studies demonstrate that cheating easily leads to a breakdown of cooperation—even if most people are initially prepared to cooperate (and even to forgive defectors at first). The existence of such cheats has been perhaps *the* classic problem for game theorists, and it has led to years of research to uncover the conditions under which cooperation can evolve while avoiding the problem of cheats. Any trait, imposed or self-emerging, that solved this problem would have conferred significant advantages.

The Natural Selection of Supernatural Punishment

One could stop here, with the argument as put forth up to this point: that religious norms, whatever their origin, may help to promote cooperation. Anthropologists and philosophers have long argued that various aspects of religion serve specific social 'functions'.³⁹ In some ways we are just continuing this tradition, simply drawing attention to a potential solution to a specific theoretical puzzle about early human cooperation. However, if religious norms really did improve cooperation, then they indirectly confer advantages to individuals in the units of biological natural selection—survival and reproductive success. As a result, religious norms promoting cooperation would have been favoured by natural selection. Many other cultural traits are subject to this process and, as John Tooby and Leda Cosmides argue, ultimately, 'human minds, human behaviour, human artifacts and human culture are all biological

35. M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge: Harvard University Press, 1965).

36. M. Ridley, *The Origins of Virtue: Human Instincts and the Origins of Cooperation* (London: Penguin, 1996).

37. Axelrod, *The Evolution of Cooperation*.

38. James Madison, 'The Structure of the Government Must Furnish the Proper Checks and Balances between the Different Departments from the New York Packet', *Federalist* 51 (8 February 1788).

39. Brian Morris, *Anthropological Studies of Religion* (Cambridge: Cambridge University Press, 1987).

phenomena'.⁴⁰ It is crucial, however, to lay out the limits here; as E.O. Wilson has written:

[G]enes do not specify elaborate conventions such as totemism, elder councils, and religious ceremonies. To the best of my knowledge no serious scientist or humanities scholar has ever suggested such a thing. Instead, complexes of gene-based epigenetic rules predispose people to invent and adopt such conventions. If the epigenetic rules are powerful enough, they cause the behaviours they affect to evolve convergently across a great many societies.⁴¹

Natural selection, then, may have favoured epigenetic predispositions receptive to religious norms of behaviour that promoted cooperation.⁴² Such norms may have conferred individual benefits even if they resulted in great risks. Wilson argues that there is 'a hereditary selective advantage to membership in a powerful group united by devout belief and purpose. Even when individuals subordinate themselves and risk death in common purpose, their genes are more likely to be transmitted to the next generation than are those of competing groups who lack equivalent resolve'.⁴³

Wilson implies here that the selection would take place at the group level. This is certainly possible, and group selection is now thought to play an important role in the evolution of numerous cultural traits.⁴⁴ However, the natural selection of religious beliefs may also occur via individual selection: as with the so-called 'green beard' effect, individuals sharing a particular cooperative trait can identify each other by signalling common attributes (such as

40. J. Tooby and L. Cosmides, 'The Psychological Foundations of Culture', in J.H. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford: Oxford University Press, 1992), pp. 20-21.

41. E.O. Wilson, *Consilience: The Unity of Knowledge* (London: Abacus, 1999), p. 184.

42. Favoured by natural selection as, more specifically, a combined product of human adaptation and culture (i.e. genetic and/or cultural selection). We should immediately preempt a predicted criticism by citing Wilson: 'In the extreme nurturist view, which has prevailed in social theory for most of the twentieth century, culture has departed from the genes and become a thing unto itself. Possessing a life of its own, growing like wildfire ignited by the strike of a match, it has acquired emergent properties no longer connected to the genetic and psychological processes that initiated it. Hence, *omnis cultura ex cultura*. All culture comes from culture.' However, we now know a number of concrete examples where 'culture has risen from the genes and forever bears their stamp. With the invention of metaphor and new meaning, it has at the same time acquired a life of its own. In order to grasp the human condition, both the genes and culture must be understood, not separately in the traditional manner of science and the humanities, but together, in recognition of the realities of human evolution.' Wilson, *Consilience*, pp. 143, 180.

43. Wilson, *Consilience*, p. 287.

44. Sober and Wilson, *Unto Others*; D.S. Wilson, 'Evolutionary Biology: Struggling to Escape Exclusively Individual Selection', *Quarterly Review of Biology* 76 (2001), pp. 199-205; D.S. Wilson and E. Sober, 'Reintroducing Group Selection to the Human Behavioural Sciences', *Behavioral and Brain Sciences* 17 (1994), pp. 585-654.

a green beard) and therefore cooperate selectively with each other and out-perform others (defectors being excluded from the benefits of mutual cooperation and its rewards). Many religions strongly promote visible symbols identifying one's commitment. Such signals are often deliberately costly, so as to make them hard to fake, and this is thought to be a crucial component of how religious groups achieve more effective cooperation.⁴⁵

We stress again that religious beliefs to punish defectors and promote cooperation in a norm-observant community would be favoured by natural selection *regardless* of the *origins* of belief in divine punishment, and *regardless* of whether divine punishment is *real* or not. In other words, if there is a real god (i.e. one that did not *initiate* out of natural selection), religious beliefs favouring cooperation—once established—could thereafter be swept along by evolutionary processes anyway.

Is This a Pessimistic View?

Some theologians have argued that a sociobiological perspective on religion, while perhaps interesting, is flawed because it implies that what is natural is what is good.⁴⁶ However, this is the so-called 'naturalistic fallacy'. Whether something has evolved or not is entirely independent of whether it is justified. The implications of natural selection (and the 'survival of the fittest' concept) for human behaviour need not be thought of as pessimistic at all. In one sense it is, because it argues that people are cooperative as a by-product of an underlying genetic self-interest (a pre-condition for natural selection). But that is focusing on an irrelevant level. Rather, natural selection is optimistic because we are, by nature, and for whatever reason, cooperative. As Matt Ridley wrote, 'if we know that an ant is altruistic only because its genes are egotistical, we still cannot deny that the ant itself is altruistic'.⁴⁷ Who cares about 'selfish' genes? What matters is that the person they make up can be a wholly cooperative altruist. Behaviour that leads to a maximization of fitness can still mean 'hard-wired' motivations for fairness and morality. Indeed, theorists continue to be flummoxed by humans' apparently irrational altruism.⁴⁸ Experimental studies reveal that humans have systematic cognitive biases towards 'fair' behaviour which often run directly counter to rational choice and their immediate interests.⁴⁹ If we are selfish in some hidden, ultimate sense, we do not know it. We lumber about helping people and cooperat-

45. Sosis, 'Religion and Intragroup Cooperation'.

46. See, for example, Roger Trigg, 'Theological Anthropology', in P. Byrne and L. Houlenden (eds.), *Companion Encyclopaedia of Theology* (London: Routledge, 1995), pp. 453-71.

47. Ridley, *The Origins of Virtue*, p. 20.

48. Johnson, Stopka and Knights, 'The Puzzle of Human Cooperation'.

49. Kagel and Roth, *The Handbook of Experimental Economics*; Henrich, 'Does Culture Matter in Economic Behaviour?'; Henrich *et al.*, 'In Search of Homo economicus'.

ing despite ourselves. A predisposition for religious beliefs and supernatural punishment would have promoted such cooperative behaviour, increasing the biological 'fitness' of those who believed.⁵⁰

Conclusions

Jupiter's secret in Sartre's play (see opening quote) encapsulates the notion that people may be better off within a community that conforms to religious norms and expects punishment for disobedience, even if believers in effect subjugate themselves to an illusory threat without 'knowing' it (i.e. they act 'as if' supernatural punishment was real and worth avoiding). The public goods achieved within such cooperative communities provide significant advantages for the individual, benefits which may outweigh any costs that are incurred in subjugation to a god or to an elite (for example, in return for taxes, peasant farmers could gain military protection from a standing army). Even if supernatural punishment has a genuine 'other-worldly' origin, as long as it stimulates public goods by promoting cooperation and deterring defectors, the system would be swept along and reinforced by natural selection. This is supported by Sosis's empirical study showing that groups with costly religious beliefs signalling commitment and loyalty outlived non-religious groups.⁵¹

Many of our social norms developed because they promoted cooperation towards public goods in the past. These norms are often driven and justified by religion. We suggest that the origins of these social norms may have spontaneously emerged in evolution as a result of the specific selective advantages of supernatural punishment. The idea that complex norm-enforcing institutions can arise spontaneously in human societies is well supported. There are many

50. One criticism of our argument is that non-believers can cheat the system simply by non-belief, thereby escaping any supernatural punishment. See E. Fehr and S. Gächter, 'The Puzzle of Human Cooperation—A Reply', *Nature* 421 (2003), p. 912. However, there are several arguments against this. First, the theory of indirect reciprocity suggests that, even if this is true, individuals who gain a reputation for following the prescribed codes of their religion and being highly cooperative within the community will gain direct benefits by the increased cooperation targeted towards them. Second, at least in hunter-gatherer societies, many factors make non-belief unlikely. For example, alternative viewpoints or scientific explanations for natural phenomena are usually absent. 'Pascal's Wager' is likely to have held significant sway in such situations: "God is or he is not." But to which side shall we incline?... Let us weigh the gain and the loss in wagering that God is. Let us estimate the two chances. If you gain, you gain all; if you lose, you lose nothing. Wager then without hesitation that he is.' Blaise Pascal, *Pensées* (1670). In addition, as Herbert Simon has argued, it is usually cheaper and more efficient to accept what other people say and follow their social norms. See H. Simon, 'A Mechanism for Social Selection of Successful Altruism', *Science* 250 (1990), pp. 1665–68.

51. Sosis, 'Religion and Intragroup Cooperation'.

examples among pre-industrial societies, which often worked far more efficiently than state-imposed ones that came later. 'Markets, exchanges and rules can develop before government or any other monopolist has defined their rules. They define their own rules, because they have been part of human nature for many millions of years.'⁵² In as much as human brains are implicated in such processes, Matt Ridley proposed the idea that: 'For St. Augustine the source of social order lay in the teachings of Christ. For Hobbes it lay in the sovereign. For Rousseau it lay in solitude. For Lenin it lay in the party. They were all wrong. The roots of social order are in our heads'.⁵³

Implications

Geographic Variation in Religion

The argument presented here offers a framework to understand geographic variation in religious traditions as adaptive solutions to local cooperation problems. It is an old idea that different religions have distinctive economic consequences,⁵⁴ and religions have been shown to directly shape and impact on political, economic and social features of their society.⁵⁵ Catholicism, for example, tends to impose higher transaction costs (all the costs involved in exchanging goods and enforcing agreements) than Anglo-Saxon religions (Weber argued that Protestantism most lends itself to capitalism). Similarly, ecological conditions in the Middle East may have contributed to the widespread practice of polygyny and for special codes of money lending.⁵⁶ Such differences may arise from specific economic pressures in the environments within which each religion formed, eliciting specific religious norms that offered the best adaptations to these challenges.⁵⁷

52. Ridley, *The Origins of Virtue*, p. 204. See also B. Benson, 'The Spontaneous Evolution of Commercial Law', *Southern Economic Journal* 55 (1989), pp. 644-61; D.W. Brown, *When Strangers Cooperate: Using Social Conventions to Govern Ourselves* (New York: The Free Press, 1994).

53. Ridley, *The Origins of Virtue*, p. 264.

54. M. Weber, *Economy and Society* (Berkeley, CA: University of California Press, 1978).

55. J.E. Lane and S. Ersson, *Culture and Politics: A Comparative Approach* (Aldershot: Ashgate, 2002).

56. J. Cartwright, *Human Evolution and Behaviour* (Cambridge: Cambridge University Press, 2000).

57. To forestall one misunderstanding of our argument, we should point out that many religions are relatively recent, so it is inconceivable that different religious groups represent different genotypes. That is not what we are suggesting at all. The argument here is that different environments create specific selective pressures which differentiate among a range of potential expressions of culture channelled by epigenetic rules shared by all humans. In other words, while the propensity for religious norms may share the same basis, they are expressed differently according to one's environment. In genetics this is known as the

Religiously Inspired Cooperation in Politics

The current decline of religion in some countries may reflect the possibility that supernatural mechanisms of norm enforcement are gradually losing salience because we now have alternative, contrived institutions to maintain cooperation by searching for, catching and punishing cheats. Transparent democratic process and public scrutiny compete with the need for supernatural incentives to promote cooperation and punish defectors. The continued strength of other religions, such as Islam, does not mean they do not have modern alternatives as well, only that those religions remain more relevant and potent in those societies, for a variety of other possible reasons.⁵⁸ It is sometimes thought that, following on from the Enlightenment writers, the 'realm of politics' has 'finally broken away from religion.'⁵⁹ However, Gilles Kepel has argued that the considerable religious revivalism in Islam, Catholicism, North American Protestantism, and Judaism witnessed in recent decades is a 'reflection of widespread and profound disquiet with modernity.'⁶⁰ This new religious approach is aimed at 'recovering a sacred foundation for the organisation of society'. He argues in particular that the aim in Islam is no longer to 'modernize Islam but to Islamize modernity'. A recent large-scale analysis suggests that despite large shifts in religious practices around the world, religion continues to be an important factor in explaining people's behaviour.⁶¹ This continues to be reflected in policy, institutional organization, the law, and the media. In many cases, it is quite clear that religious codes meant to enforce cooperation and punish defectors still have considerable currency in modern legal and political discourse (for example, swearing on the Bible to tell the truth in court, and the appeal to religious duty and dangers of evil in US foreign policy).⁶²

Summary

It seems a compelling possibility that religious belief in supernatural punishment greatly aided the evolution of cooperation in humans. This mechanism has four major selective advantages that evade the classic public goods problems troubling current theoretical work: First, defectors are automatically punished. Second, there is no second-order free-rider problem (God does

'reaction norm' of a trait, which signifies its malleability among different populations. Culture may shape it too, of course.

58. One of which is the imposition of religious norms by force, via religious police, etc.

59. Gilles Kepel, *The Revenge of God: The Resurgence of Islam, Christianity and Judaism in the Modern World* (University Park, PA: Pennsylvania State University Press, 1994), p. 2.

60. Kepel, *The Revenge of God*, p. 2.

61. Lane and Ersson, *Culture and Politics*.

62. See, for example, William Martin, *With God on Our Side: The Rise of the Religious Right in America* (New York: Broadway Books, 1997).

everything). Third, no one has to be a vigilante and risk reprisals. Fourth, it is an extremely efficient deterrent because free-riders are always caught, even before the act, because gods are omnipresent and omniscient. Evil thoughts are punishable offences as well as evil acts (for example, Mt. 5:28: 'whosoever looketh on a woman to lust after her hath committed adultery with her already in his heart'). If the argument presented here remains unconvincing, then at least consider Wilson's challenging point that: 'the human mind evolved to believe in the gods. It did not evolve to believe in biology.'⁶³ As far back as we can see into human history, and as much as we know of the world's societies, the fear of God or supernatural spirits has proven to be a tremendously successful method of coercing people to cooperate, even (via religious or political leaders acting in their name) among *non*-believers. Whatever its origin, one cannot deny that supernatural punishment offers an excellent solution to the theoretical problems that so endanger human cooperation. As Voltaire said, 'if God did not exist, it would be necessary to invent him'.⁶⁴

Acknowledgments: We are grateful for comments, criticism and discussion leading to this paper from Luis Zaballa, Jan-Erik Lane, Pavel Stopka, Stephen Knights, Dominic Tierney, Paul Johnson, Pierre Allan, Stephen Peter Rosen, Richard Wrangham, Nick Brown, Richard Sosis and Gabriella de la Rosa. DDPJ thanks the Swiss Federal Government for funding.

Dominic Johnson is a Science Fellow at the Center for International Security and Cooperation at Stanford University. He was formerly a research fellow at Harvard, after completing his doctorate in evolutionary biology at Oxford and a Masters degree in political science at Geneva University. His research focuses on the origins of cooperation, conflict over resources, and biases in decision-making.

Oliver Krüger is a Royal Society Research Fellow in the Department of Zoology, University of Cambridge. For his university studies and PhD, he went to the Universities of Bielefeld, Oxford and Cambridge. His research interests are in the field of life history strategies, the way evolution shapes an organism's allocation of resources to growth, reproduction and partner choice.

63. Wilson, *Consilience*, p. 292.

64. Voltaire, 'A l'Auteur du Livre des Trois Imposteurs', *Épîtres* 96 (1770).

BIBLIOGRAPHY

- Axelrod, R., *The Evolution of Cooperation* (London: Penguin, 1984).
- The Complexity of Cooperation: Agent-based Models of Competition and Collaboration* (Princeton: Princeton University Press, 1997).
- Barkow, J.H., L. Cosmides and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford: Oxford University Press, 1992).
- Benson, B., 'The Spontaneous Evolution of Commercial Law', *Southern Economic Journal* 55 (1989), pp. 644-61.
- Bernstein, A.E., *The Formation of Hell: Death and Retribution in the Ancient and Early Christian Worlds* (Cornell: Cornell University Press, 1993).
- Boyd, R., and P.J. Richardson, 'Why People Punish Defectors: Weak Conformist Transmission Can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas', *Journal of Theoretical Biology* 208 (2001), pp. 79-89.
- Brown, D.W., *When Strangers Cooperate: Using Social Conventions to Govern Ourselves* (New York: The Free Press, 1994).
- Byrne, P., and L. Houlden (eds.), *Companion Encyclopaedia of Theology* (London: Routledge, 1995).
- Cartwright, J., *Human Evolution and Behaviour* (Cambridge: Cambridge University Press, 2000).
- Clutton-Brock, T.H., and G.A. Parker, 'Punishment in Animal Societies', *Nature* 373 (1995), pp. 209-216.
- Earhart, H.B. (ed), *Religious Traditions of the World* (New York: HarperCollins, 1993).
- Fehr, E., and S. Gächter, 'Cooperation and Punishment in Public Goods Experiments', *American Economic Review* 90 (2000), pp. 980-94.
- 'Altruistic Punishment in Humans', *Nature* 415 (2002), pp. 137-40.
- 'The Puzzle of Human Cooperation—A Reply', *Nature* 421 (2003), p. 912.
- Gintis, H., E. Smith and S. Bowles, 'Costly Signalling and Cooperation', *Journal of Theoretical Biology* 213 (2001).
- Hackathorn, D.D., 'Collective Action and the Second-order Free-rider Problem', *Rational Society* 1 (1989), pp. 78-100.
- Hamilton, W.D., 'The Genetical Evolution of Social Behaviour, I & II', *Journal of Theoretical Biology* 7 (1964).
- Harrison, E., G. Bromiley and C. Henry (eds.), *Wycliffe Dictionary of Theology* (Peabody, MA: Hendrickson, 1960).
- Hashmi, S.H., 'Is there an Islamic Ethic of Humanitarian Intervention?', *Ethics and International Affairs* 7 (1993), pp. 55-73.
- Henrich, J., 'Does Culture Matter in Economic Behaviour? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon', *American Economic Review* 90 (September 2000), pp. 973-79.
- Henrich, J., *et al.*, 'In Search of Homo economicus: Behavioural Experiments in 15 Small-scale Societies', *American Economic Review* 91.2 (2001), pp. 73-78.
- Johnson, D.D.P., P. Stopka and S. Knights, 'The Puzzle of Human Cooperation', *Nature* 421 (2003), pp. 911-12.
- Kagel, J.H., and A.E. Roth (eds.), *The Handbook of Experimental Economics* (Princeton: Princeton University Press, 1995).
- Kepel, Gilles, *The Revenge of God: The Resurgence of Islam, Christianity and Judaism in the Modern World* (University Park, PA: Pennsylvania State University Press, 1994).

- Lane, J.E., and S. Ersson, *Culture and Politics: A Comparative Approach* (Aldershot: Ashgate, 2002).
- Leakey, R., and R. Lewin, *Origins Reconsidered: In Search of What Makes Us Human* (London: Anchor, 1993).
- Locke, John, *Treatise of Civil Government and a Letter Concerning Toleration* (ed. C. Sherman; New York: Appleton-Century, 1937).
- Madison, James, 'The Structure of the Government Must Furnish the Proper Checks and Balances between the Different Departments from the New York Packet', *Federalist* 51 (February 1788).
- Martin, William, *With God on Our Side: The Rise of the Religious Right in America* (New York: Broadway Books, 1997).
- Morris, Brian, *Anthropological Studies of Religion* (Cambridge: Cambridge University Press, 1987).
- Murdock, G.P., 'The Common Denominator of Cultures', in *idem* (ed.), *The Science of Man in the World Crisis* (New York: Columbia University Press, 1945).
- Olson, M., *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge: Harvard University Press, 1965).
- Ostrom, E., J. Walker and R. Gardner, 'Covenants With and Without a Sword: Self Governance Is Possible', *American Political Science Review* 86 (1992), pp. 404-417.
- Ridley, M., *The Origins of Virtue: Human Instincts and the Origins of Cooperation* (London: Penguin, 1996).
- Sartre, Jean-Paul, *Les Mouches* (Paris: Gallimard, 1947).
- Schelling, T.C., *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960).
- Sigmund, K., C. Hauert and M. Nowak, 'Reward and Punishment', *Proceedings of the National Academy of Sciences USA* 98 (2001), pp. 10757-61.
- Simon, H., 'A Mechanism for Social Selection of Successful Altruism', *Science* 250 (1990), pp. 1665-68.
- Sober, E., and D.S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behaviour* (Cambridge: Harvard University Press, 1998).
- Sosis, R., 'Religion and Intragroup Cooperation: Preliminary Results of a Comparative Analysis of Utopian Communities', *Cross-Cultural Research* 34 (2000), pp. 71-88.
- Thomas, W.L., and D.S. Thomas, *The Child in America: Behaviour Problems and Programs* (New York: Alfred A. Knopf, 1928).
- Tooby, J., and L. Cosmides, 'The Psychological Foundations of Culture', in J.H. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford: Oxford University Press, 1992).
- Trigg, Roger, 'Theological Anthropology', in P. Byrne and L. Houlden (eds.), *Companion Encyclopaedia of Theology* (London: Routledge, 1995).
- Trivers, R.L., 'The Evolution of Reciprocal Altruism', *Quarterly Review of Biology* 46 (1971).
- Voltaire, 'A l'Auteur du Livre des Trois Imposteurs', *Épîtres* 96 (1770).
- Weber, M., *Economy and Society* (Berkeley, CA: University of California Press, 1978).
- *The Sociology of Religion* (Berkeley: University of California Press, 1978).
- Wedekind, C., and M. Milinski, 'Cooperation through Image Scoring in Humans', *Science* 288 (2000).
- Welch, D.A., *Justice and the Genesis of War* (Cambridge: Cambridge University Press, 1993).
- Wilson, D.S., 'Evolutionary Biology: Struggling to Escape Exclusively Individual Selection', *Quarterly Review of Biology* 76 (2001), pp. 199-205.
- Wilson, D.S., and E. Sober, 'Reintroducing Group Selection to the Human Behavioural Sciences', *Behavioral and Brain Sciences* 17 (1994), pp. 585-654.
- Wilson, E.O., *Consilience: The Unity of Knowledge* (London: Abacus, 1999).