Wrath of God: religious primes and punishment

Ryan McKay\textsuperscript{1,2,3,4,*}, Charles Efferson\textsuperscript{1,2}, Harvey Whitehouse\textsuperscript{3} and Ernst Fehr\textsuperscript{1,2}

\textsuperscript{1}Institute for Empirical Research in Economics, and \textsuperscript{2}Laboratory for Social and Neural Systems Research, University of Zürich, Zürich, Switzerland
\textsuperscript{3}Centre for Anthropology and Mind, University of Oxford, Oxford, UK
\textsuperscript{4}Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

Recent evidence indicates that priming participants with religious concepts promotes prosocial sharing behaviour. In the present study, we investigated whether religious priming also promotes the costly punishment of unfair behaviour. A total of 304 participants played a punishment game. Before the punishment stage began, participants were subliminally primed with religion primes, secular punishment primes or control primes. We found that religious primes strongly increased the costly punishment of unfair behaviours for a subset of our participants—those who had previously donated to a religious organization. We discuss two proximate mechanisms potentially underpinning this effect. The first is a ‘supernatural watcher’ mechanism, whereby religious participants punish unfair behaviours when primed because they sense that not doing so will enrage or disappoint an observing supernatural agent. The second is a ‘behavioural priming’ mechanism, whereby religious primes activate cultural norms pertaining to fairness and its enforcement and occasion behaviour consistent with those norms. We conclude that our results are consistent with dual inheritance proposals about religion and cooperation, whereby religions harness the byproducts of genetically inherited cognitive mechanisms in ways that enhance the survival prospects of their adherents.

Keywords: religion; supernatural agency; subliminal priming; fairness; punishment; cooperation

1. INTRODUCTION

...The \textsc{lord} is a jealous God, filled with vengeance and wrath...

\textit{(Nahum 1:2)}

Religion carries formidable epistemic, metabolic and material costs [1–3]. Religious believers must maintain and compartmentalize beliefs that are extravagantly at variance with intuitive conceptions of reality. Religious rituals, moreover, are often physically taxing and painful, and frequently require the sacrifice of precious resources. Given such costs, some evolutionary theorists argue that religion must provide, or in the ancestral past must have provided, countervailing adaptive benefits (e.g. [1–7]; cf. [8,9]). Perhaps the most influential of such proposals is that religion is a cultural variant that confers a selective advantage at the group level by virtue of the fact that it secures and promotes cooperative behaviour within the group [6,7]. This proposal arguably solves not one but two thorny evolutionary puzzles: the puzzle of religion and the puzzle of human cooperation.

The nature and extent of human cooperation is unique in the animal kingdom [10,11]. Human societies are based on large-scale cooperation between genetically unrelated individuals. Cooperation is frequent in non-repeated interactions, even when reputational gains are small or absent. Cooperation and other prosocial behaviours will in many situations be sustained by preferences for \textit{fairness}, or by a cultural norm of fairness [10]. Such preferences are evidenced by the behaviour of participants in anonymous, one-shot economic games, many of who nominate fair outcomes even when such outcomes are disadvantageous with respect to their material self-interest [10,12]. Humans, moreover, reward others who behave fairly and impose sanctions on those who fail to do so [10].

One potential means of implementing fairness norms is via culturally postulated supernatural agents [9], in particular ‘full-access strategic agents’ such as omnipotent, omniscient, moralizing gods [8,13]. Individuals who believe that behavioural norms are policed by an all-knowing supernatural agent with the power and inclination to inflict terrible retribution for norm violations will have a strong incentive to comply with those norms. Some authors, therefore (e.g. [7,14–16]), have suggested that belief in supernatural punishment confers a selective advantage by promoting prosocial behaviour.

Recent evidence from studies employing priming paradigms is consistent with this proposal. In a seminal study, Shariff & Norenzayan [17] used a scrambled sentence task to prime religious concepts, and found that participants primed in this fashion gave significantly more money in a subsequent (anonymous, one-shot) Dictator game than did control participants. Similar results have been found in other recent priming studies. For example, relative to control participants, participants primed with religious or supernatural concepts have been found to
In the second stage of the game, player B was informed of the two options that were available to player A in the first stage, but did not learn the specific choice that player A made. Instead, we used the ‘strategy method’ in order to maximize the amount of statistical data gathered: player B was, for each option, given the opportunity to spend points out of her allocation share in order to reduce player A’s payoff in that case—i.e. to punish player A. In each case, the choice was binding, provided that the relevant option was actually chosen by player A. Previous work has shown that participants’ qualitative behavioural patterns are unaffected by the use of this method as opposed to the ‘direct-response’ method, in which player B learns the specific choice made by player A and then chooses a response [25,26]. Player B could spend a maximum of 50 points (and minimum of 0) to punish player A, i.e. player B entered a number between 0 and 50 for each of the two options. A 1 : 3 punishment ratio was employed, such that each point spent by player B reduced the payoff of player A by three points—provided that the relevant choice was actually made by player A. Hence, if player A chose the allocation \((x_A,x_B)\) and player B punished her with \(0 \leq p \leq 50\) points for that choice, player A’s payoff was \(x_A - 3p\) and player B’s payoff was \(x_B - p\).

(c) Priming episode and visibility check
The priming episode followed the first stage of the punishment game. There were four between-subjects priming treatments:

(i) religion (primes: divine, holy, pious, religious);
(ii) punishment (primes: revenge, punish, penalty, retribution);
(iii) religion–punishment (primes: divine, revenge, pious, punish); and
(iv) control (primes: northeast, acoustic, tractor, cartoon).

The priming episode comprised 20 priming trials. The sequence of events for each trial was as follows (figure 1): fixation point for 500 ms, forward mask for 500 ms, prime for 40 ms then backward mask for 500 ms. As soon as the priming episode concluded, the second stage of the punishment game began, and once this stage was complete participants underwent a systematic test for prime visibility. We excluded any participants who performed significantly above chance on this test.

(d) Religion questionnaire
In addition to requesting religious affiliation, our religion questionnaire included a series of items answered on a five-point Likert scale (1 = strongly disagree; 5 = strongly agree; see table 1 for a list of these items) followed by...
3. RESULTS

We ran separate analyses for two dependent variables:

(i) punishment of the unfair choice (590, 60) (table 2) and
(ii) punishment of the fair choice (150, 150) (table 3).

In each case, we analysed the actual strength (amount) of punishment rather than simply whether punishment occurred or not at any strength.

There were no significant main effects of our experimental primes on punishment of the unfair (table 2; model 1) or fair (table 3; model 1) choices, although the effect of punishment primes was marginally significant (and positive) for punishment of the unfair choice (see coefficient for ‘punishment prime’ in table 2; model 1). For punishment of the unfair choice, however, there was a highly significant interaction between religious donations and the religion priming treatment (see coefficient for ‘religious donations × rel’ in table 2, model 1). Neither religious donations nor religious priming had effects in isolation (see respective coefficients for ‘religious donations’ and ‘religion prime’ in table 2, model 2), but when both were present they jointly increased punishment of the unfair choice by 0.84 of a standard deviation, \( p = 0.001 \). Apart from a significant main effect of gender for punishment of the unfair choice (see coefficient for ‘female’ in table 2, model 1: in line with previous research, females were less punitive; (e.g. [27])), no other effects were significant for either type of punishment.

4. DISCUSSION

Humans are subject to strong cultural norms of fairness. A substantial proportion of participants in anonymous, one-shot economic games nominate fair outcomes even when such outcomes are to their material disadvantage [10,12]. Humans, moreover, incur costs to reward others who behave fairly and to impose sanctions on those who behave unfairly [11]. Recent studies indicate that priming participants with religious concepts promotes prosocial behaviour (e.g. [17,19]). Our aim in the present experiment was to investigate whether religious priming would also promote the costly punishment of unfair behaviour.

Across all participants, our results indicate a negative answer to this question: there was no main effect of religious primes, whether alone or in combination with punishment primes, on punishment behaviour. The only suggestion of a main effect for priming treatment was for punishment primes alone, which (perhaps unsurprisingly) tended to increase punishment of unfair choices. However, religious primes did strongly increase the costly punishment of unfair behaviour for a subset of our participants—those who had previously donated to a religious organization.

How are we to account for these results? In line with Shariff & Norenzayan [17], we consider two possible proximate explanations. The first is that religious primes activate the notion that one’s behaviour is observed by a supernatural agent. In this case primed participants punish unfair behaviours because they sense that not doing so will damage their standing in the eyes of a supernatural agent. Recent studies suggest that even very subtle cues that one is being watched, such as stylized eyespots on a computer screen, can affect giving behaviour (e.g. [28,29]; cf. [30]). Some authors have suggested that such cues match the input conditions for evolved mental mechanisms that detect when one’s behaviour is observed [28]. Religious primes might likewise function as input for these mechanisms [17].

The second possibility is a behavioural priming explanation, whereby religious primes activate cultural norms pertaining to fairness and its enforcement and occasion behaviour consistent with those norms. This explanation is consistent with the evidence that the activation of conceptual representations increases the likelihood of behaviours consistent with those representations (e.g. [31]). Thus, much as participants walk more slowly down a length of corridor when the concept ‘elderly’ is primed [31], priming words that are semantically associated with fairness may lead participants to punish unfair behaviours simply by virtue of that semantic connection [22]. Bargh et al. [32] found that participants primed with cooperation-related words (e.g. fair, share) were less selfish in a subsequent resource-management game, and Shariff & Norenzayan [17] found that priming with secular-moral words (e.g. courts, contract) had a similar effect to that of religious primes on fair allocations in a subsequent Dictator game.

Although they acknowledge that the two mechanisms above need not be mutually exclusive, Shariff & Norenzayan [17] favour the ‘supernatural watcher’ account (cf. [22]). Norenzayan et al. [33] argue that behavioural priming effects are ‘typically impervious to prior explicit beliefs or attitudes’ (p. 532). If this is true, then one would not expect the effects of religious primes to be mediated by individual religiosity if those effects are attributable to behavioural priming. Norenzayan et al. [33] also describe a recent series of studies which found that religious primes caused an increase in public self-awareness, which is ‘characterized by attentiveness to those features of one’s self that are presented to others’ ([34], p. 366) and so directly linked to sensitivity about being observed.

Regarding the former point, it seems to us that the effect of activating a certain set of cultural norms might well be stronger for those who have internalized those norms. With respect to our experiment, it seems plausible.
that the behavioural norms of religious institutions are more strongly represented in the minds of individuals who financially support those institutions, and thus more susceptible to activation by relevant primes. Nevertheless, we agree with Norenzayan et al. [33] that multiple psychological mechanisms may be operative and even mutually reinforcing. If an individual believes that in order to avoid punishment herself she needs both to adhere to and to uphold cultural norms of fairness, then religious primes may affect punishment behaviours both by evoking a sense of being observed and by directly activating the relevant norms. Future work might profitably investigate these possibilities.

The accounts we have considered above pertain to proximate explanation. In terms of ultimate evolutionary explanation, our results are consistent with dual inheritance proposals about religion and cooperation4. A number of authors (e.g. [35–38]) have suggested that the human proclivity for acquiring and transmitting supernatural agent concepts is an incidental byproduct of cognitive mechanisms genetically adapted for other purposes. Others (e.g. [33,39–42]) have argued that religions are cultural systems that exploit such byproducts to adaptive effect. If, as our results indicate, the activation of supernatural agent concepts promotes the enforcement of cultural norms of fairness, and if such norms sustain cooperative behaviours within the group, then religions that harness such concepts will enhance the survival prospects of their vectors, thereby contributing to their own survival.

This experiment was approved by (and conducted at) the Institute for Empirical Research in Economics, University of Zurich.

This research was supported by the European Commission’s Sixth Framework Programme (‘Explaining Religion’), the Swiss National Science Foundation (grant no. 105312-114107) and the University of Zurich Research Priority Program ‘Foundations of Human Social Behaviour—

Table 2. Models 1 and 2 for punishment of the unfair choice. (Predictor variables include age and a dummy variable denoting female gender; a composite of the 11 Likert items from the religion questionnaire, representing the average of responses to these items; a dummy variable for religious donations; and dummies for the three experimental priming treatments. Each of the priming treatment dummy variables is also interacted with the Likert composite and with the religious donations dummy.)

<table>
<thead>
<tr>
<th>variable</th>
<th>model 1</th>
<th>model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>s.e.</td>
</tr>
<tr>
<td>intercept</td>
<td>6.966</td>
<td>1.966</td>
</tr>
<tr>
<td>age</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>female</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Likert composite</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religion prime</td>
<td>-0.170</td>
<td>2.844</td>
</tr>
<tr>
<td>punishment prime</td>
<td>5.275</td>
<td>2.793</td>
</tr>
<tr>
<td>religion–punishment prime</td>
<td>1.960</td>
<td>2.844</td>
</tr>
<tr>
<td>Likert composite × rel</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Likert composite × pun</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Likert composite × rel–pun</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations × rel</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations × pun</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations × rel–pun</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*p = 0.05.

**p = 0.01.

Table 3. Models 1 and 2 for punishment of the fair choice. (Predictor variables are as per table 2.)

<table>
<thead>
<tr>
<th>variable</th>
<th>model 1</th>
<th>model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>s.e.</td>
</tr>
<tr>
<td>intercept</td>
<td>1.610</td>
<td>1.314</td>
</tr>
<tr>
<td>age</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>female</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Likert composite</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religion prime</td>
<td>0.760</td>
<td>1.901</td>
</tr>
<tr>
<td>punishment prime</td>
<td>3.079</td>
<td>1.867</td>
</tr>
<tr>
<td>religion–punishment prime</td>
<td>2.131</td>
<td>1.901</td>
</tr>
<tr>
<td>Likert composite × rel</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Likert composite × pun</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Likert composite × rel–pun</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations × rel</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations × pun</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>religious donations × rel–pun</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
were university students in Zurich, a note of caution about the generalizability of our findings is in order. Given that most of our participants were university students, particularly in Western societies, are frequent outliers on psychological measures. When we re-analyse with normalized continuous variables, the coefficient for ‘religious donations x rel’ is 0.84.

1 Recently, Henrich et al. [43] documented evidence that university students, particularly in Western societies, are frequent outliers on many psychological measures. Given that most of our participants were university students in Zurich, a note of caution about the generalizability of our findings is in order.

REFERENCES


