ELSEVIER

Evolution and Human Behavior 26 (2005) 375–387

Evolution and Human Behavior

Altruistic punishing and helping differ in sensitivity to relatedness, friendship, and future interactions

Rick O'Gorman[a], David Sloan Wilson[a,b,*], Ralph R. Miller[c]

[a]*Department of Biological Sciences, Binghamton University, Binghamton, NY 13902-6000, USA*
[b]*Department of Anthropology, Binghamton University, Binghamton, NY 13902-6000, USA*
[c]*Department of Psychology, Binghamton University, Binghamton, NY 13902-6000, USA*

Initial receipt 26 January 2004; final revision received 17 December 2004

Abstract

Altruism is behaviorally defined as an act that benefits others at the expense of the actor. Altruism is usually associated with helping others in need, but it can also take place in the context of punishment. People who help to maintain cooperation by punishing cheaters are benefiting others at their own expense as surely as if they performed acts of overt helping. The proximate psychological mechanisms that motivate altruistic helping and altruistic punishment are almost certainly different from each other (e.g., empathy vs. moralistic anger). We present two studies suggesting that the impulse to altruistically help and altruistically punish differ in their sensitivity to information regarding genetic relatedness and probability of future interactions. This interesting empirical result is relevant to the interpretation of altruistic punishment as an evolved adaptation versus a byproduct of modern environments, and to the evolution of psychological traits associated with morality.
© 2005 Published by Elsevier Inc.

Keywords: Altruism; Altruistic punishment; Altruistic helping; Kin selection; Reciprocity; Morality

* Corresponding author. Department of Biological Sciences, Binghamton University, Binghamton, NY 13902-6000, USA. Tel.: +1 607 777 4393; fax: +1 607 777 6521.
E-mail address: dwilson@binghamton.edu (D.S. Wilson).

1090-5138/05/$ – see front matter © 2005 Published by Elsevier Inc.
doi:10.1016/j.evolhumbehav.2004.12.006

1. Introduction

Altruism is defined by evolutionary biologists as an act that increases the fitness of others while decreasing the fitness of the actor. This definition makes no reference to the psychological mechanisms that motivate the act, which might or might not count as altruistic as psychologists and philosophers use the term. Fitness effects and psychological mechanisms can be understood and related to each other in terms of the evolutionary distinction between ultimate and proximate causation, as reviewed by Sober and Wilson (1998).

Altruism has traditionally been studied in the context of helping, but punishment can also qualify as altruistic in terms of its fitness effects. Consider the simultaneous evolution of two traits: (1) an altruistic act such as providing a public good versus free-riding, and (2) punishing free riders versus failing to punish. Nonpunishers benefit from the policing efforts of the punishers in just the same way that free riders benefit from the public good provided by the altruists. Theoretical models show that altruistic punishment can favor the evolution of other altruistic acts that would not evolve in the absence of punishment (Boyd, Gintis, Bowles, & Richerson, 2003; Boyd & Richerson, 1992; Gintis, 2000; Henrich & Boyd, 2001). Nevertheless, altruism supported by punishment does not necessarily evolve to fixation. Instead, a behavioral polymorphism often results, which includes public good providers, free riders, punishers, and nonpunishers.

Strong experimental evidence exists for this behavioral mix in humans, thanks largely to the efforts of economists (Fehr, Fischbeher, & Gachter, 2002; Fehr & Gachter, 2000, 2002; Ostrom, Gardner, & Walker, 1994). In games where a public good can be provided at private expense, individuals differ in their tendency to free ride. Once the more generous members of the group realize that they are being exploited by free riders, they tend to withhold their altruism, resulting in the complete absence of public good provision. When the opportunity to punish free riders at private expense is added, some individuals elect to punish, in addition to providing the public good. A sufficient number of punishers make free-riding disadvantageous, and public good provision rises to near maximum levels. Punishment can be regarded as self-interested, despite its private cost, if the punisher benefits from increased public good provision over the long term. However, numerous experiments have been performed in which the return benefits of punishment are rigorously excluded (Fehr & Gachter, 2000, 2002). These experiments show that some individuals punish only when it is in their perceived self-interest, but a sizeable fraction continues to punish in the complete absence of return benefits. These experiments unequivocally demonstrate the existence of altruistic punishment in humans at the behavioral level. The pronounced individual differences are also in accord with the aforementioned theoretical models.

While the existence of altruistic punishment in humans is well established, its interpretation has become the subject of a vigorous debate. One possibility is that altruistic punishment is an adaptation that evolved because "groups with a high fraction of altruistic punishers would have sustained cooperation more successfully than groups with fewer punishers, and so would have prevailed over them" (Bowles & Gintis, 2002, p. 128).

Another possibility is that altruistic punishment is maladaptive in a modern context and only makes sense in relation to the human ancestral environment. According to this mismatch hypothesis, people lack the cognitive adaptations for behaving appropriately in the context of the experiments because social interactions invariably took place among genetic relatives or nonrelatives with a high likelihood of future interactions (Johnson, Stopka, & Knights, 2003).

In this paper, we use fictional scenarios to explore the effects of information regarding genetic relatedness, friendship, and potential for future interactions in situations that invoke altruistic punishment and altruistic helping. We conducted two experiments, using fictional scenarios that were similar to the aforementioned games involving actual interactions. The first experiment was designed to test the effects of information regarding genetic relatedness, friendship, and potential for future interactions on the desire to punish norm violations. The second experiment simplified and changed some details of the transgression scenario and added an altruistic helping scenario. The advantage of this methodology is that the elements of the scenarios can be systematically varied and presented to large numbers of people in a way that would be difficult or impossible to stage with actual interactions. Responses to fictional scenarios provide important insights into psychological mechanisms, even when they do not correspond directly to responses to actual interactions. In our case, we demonstrate a remarkable difference between altruistic helping and altruistic punishment in their sensitivity to information regarding genetic relatedness, friendship, and potential for future interactions. This difference is relevant to the debate over altruistic punishment and more generally to the evolution of psychological mechanisms associated with morality.

2. Experiment 1

2.1. Methods

Four hundred seventy undergraduates (190 males, 279 females; 1 not reported; ages between 16 and 35 years, with a mean of 18.8) from an introductory Psychology course at SUNY-Binghamton completed the present study as part of a mass testing session for course credit.

Participants were presented with a questionnaire that asked them to imagine themselves as members of an investment club who pool individual contributions of US$1000 each to play the stock market. Information regarding friendship and genetic relatedness was manipulated by describing the members as cousins, friends, or strangers prior to the formation of the club. The probability of future interaction was manipulated by stating that everyone in the club remained in the same town and were likely to interact again, that the cheater had moved to another town and was unlikely to interact with any other member of the club, or that the subject had moved to another town and was unlikely to interact with the cheater, who nevertheless was likely to interact with other members of the club. These conditions were crossed in a 3×3 between-group factorial design, yielding nine treatments.

Participants are told that the investments triple in value, but one member cheats and takes US$2160 more than he deserves (see Appendix A for one version of the scenario). The questionnaire then directs participants to respond to the following questions:

(1) How angry would you feel toward this person? Please answer on a scale from 1 (*not at all angry*) to 9 (*extremely angry*)_____
(2) How much would you like to punish this person? Please answer on a scale from 1 (*no interest in punishing*) to 9 (*extremely interested in punishing*)_____
(3) Although punishment can take many forms, if you think of it as a dollar amount, how much do you think this person should pay for what he or she did? US$_____
(4) Punishing this person can take many forms, but if you think of it as a dollar amount, what is the most you would be willing to pay to punish this person? US$_____

2.2. Results

Considering all treatments together, participants responded to the scenario with a high average degree of anger (M=7.4 on a nine-point scale, S.D.=1.7) and desire to punish the transgressor (M=6.1, S.D.=2.3). On average, they felt that the transgressor should pay back approximately what he stole (US$2089.4, S.D.=2576.8), and they were prepared to pay a mean of US$421.3 (S.D.=2140.1) to punish the transgressor. However, there were no statistically significant differences among the treatments ($p>.2$). The variables of genetic relatedness, friendship, and potential for future interactions appeared to be psychologically irrelevant. There was also no main effect or interaction effects for sex, although it approached significance for a main effect on desire to punish the transgressor ($p=.083$; $M_{\text{Males}}=6.4$, $\text{S.D.}_{\text{Males}}=2.2$; $M_{\text{Females}}=5.9$, $\text{S.D.}_{\text{Females}}=2.4$).

Another important result concerns the distribution of scores (see Fig. 1), in addition to their mean values. The self-report of anger and desire to punish were so strong that the distributions were truncated at the high end. In fact, the most common response for anger was the maximum value of nine. The distribution for the amount that the transgressor should pay includes a large spike corresponding to the amount that he stole, with other participants indicating that he should pay more (retribution) or paradoxically less than he stole. The distribution for the amount that the participant would be willing to pay to punish the transgressor shows that a large fraction of participants are willing to pay nothing or very little, while others are willing to pay a great deal. Taken together, these distributions suggest that, while most participants are very angry and would like to see the transgressor punished, only some are willing to altruistically punish.

3. Experiment 2

3.1. Methods

Four hundred six undergraduates (159 males, 203 females; 44 not reported; ages between 17 and 33 years, with a mean of 18.9) from an introductory Psychology course

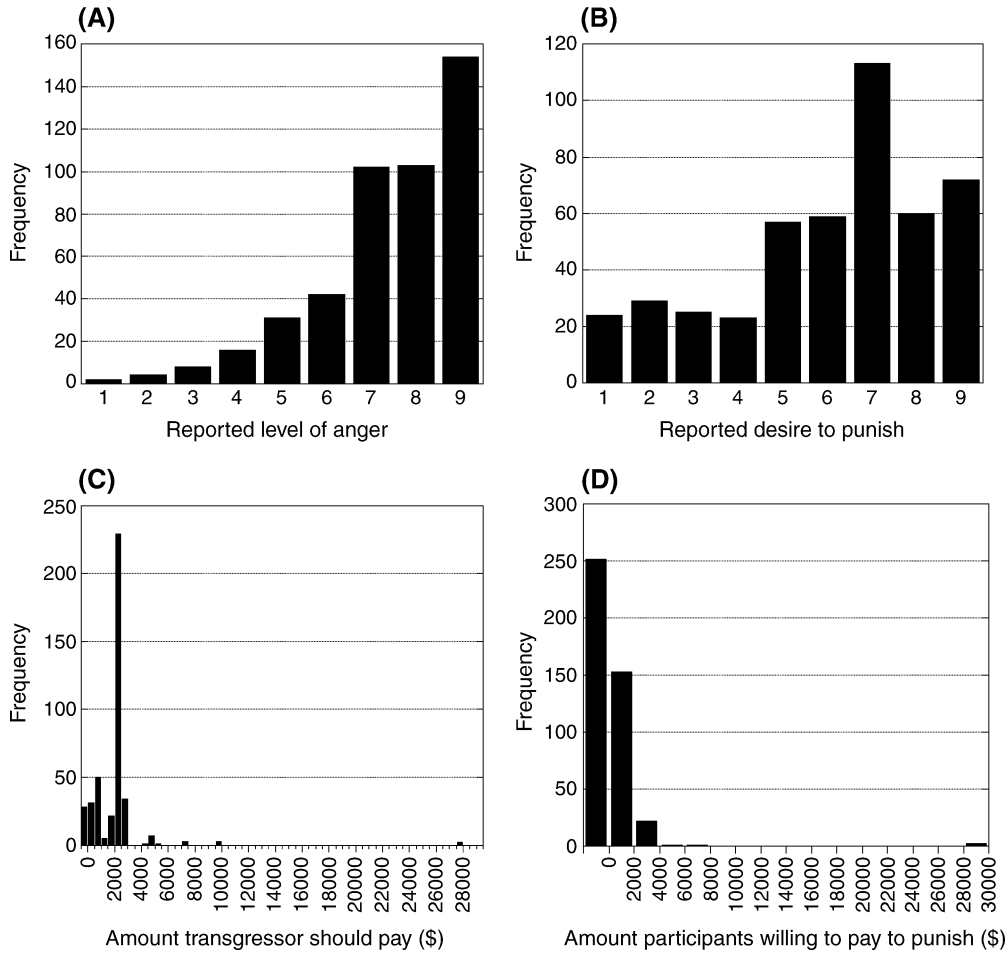

Fig. 1. Histograms of participants' responses to the fictional scenarios in Experiment 1 [*n* is 463, 463, 417 and 431 for (A), (B), (C), and (D), respectively]. Panel (A) shows the distribution of responses for reported levels of anger, showing that a majority of participants reported a high level of anger. Panel (B) shows that a substantial number of participants also had a high desire to punish the transgressor. Panel (C) shows the amount of money that participants thought the cheater should repay, with intervals of US$500 (the *x* axis labels show the upper limit of the corresponding interval). Panel (D) shows the amount of money that participants reported being willing to pay to punish the transgressor, with intervals of US$2000 (the *x* axis labels show the upper limit of the corresponding interval). The data show a dissociation between participants' reported emotions and their willingness to punish.

at SUNY-Binghamton completed the present study as part of a mass testing session for course credit.

The materials and procedure were similar to those for Experiment 1. A second punishment scenario was added, in which a fellow group member proposes going to the socially sanctioned mechanism of small claims court to seek redress, in addition to punishing the transgressor in an unspecified fashion, as per Experiment 1. This version was added to determine whether the results from our first experiment were due to a lack of a socially

acceptable (normative) means to punish the cheater. We also added an altruistic helping scenario, in which one investing member had anticipated a profit to pay for emergency medical bills and the participants were asked how much they would be willing to contribute toward helping.

Information regarding friendship and genetic relatedness was manipulated as in the first experiment, but the "friends" treatment was dropped to limit the total number of treatments to accommodate the socially sanctioned punishment scenario and the altruistic helping scenario while maintaining adequate sample sizes. The probability of future interaction was also manipulated as in the first experiment, with the treatment in which the participant had moved to another town dropped, again to limit the number of treatments, thus producing a $3\times2\times2$ between-group factorial design, yielding 12 treatments.

Participants were presented with a questionnaire that asked them to imagine themselves as members of an investment club who pool individual contributions of US$1000 each to play the stock market. In the transgression scenario, the investments merely break even rather than making a profit, except for the amount stolen by the transgressor, US$200. In the helping scenario, the investments break even.

In the transgression scenario, participants were presented with the same questions as used in Experiment 1. Participants in the helping scenario were presented with the following questions:

(1) How sorry would you feel toward this person? Please answer on a scale from 1 (*not at all sorry*) to 9 (*extremely sorry*)_____
(2) How much would you like to help this person? Please answer on a scale from 1 (*no interest in helping*) to 9 (*extremely interested in helping*)_____
(3) What is the most you would be willing to give to help this person? US$_____
(4) Some of the other group members want to help the person, but others seem unwilling to help if it costs them anything. How angry would you feel toward those who do not want to help? Please answer on a scale from 1 (*not at all angry*) to 9 (*extremely angry*)_____

3.2. Results

The results of the second experiment confirmed the results of the first experiment in all essential details: Participants in the punishment treatments indicated strong anger at the transgression (M=6.4, S.D.=2.2) and a moderate to high desire to punish (M=5.4, S.D.=2.4) but were insensitive to information regarding genetic relatedness or potential for future interactions ($p>.16$ for both main effects). Most wanted the transgressor to pay back what had been stolen (52.1%), and many were unwilling to pay to punish the transgressor (48.1%), despite their anger (see Fig. 2). Providing a socially sanctioned mechanism for punishment (small claims court) significantly decreased the average level of anger [M=5.9, S.D.=2.2 for the treatments with small claims court option vs. M=6.7, S.D.=2.2; $F(1,247)=8.58$, $p=.004$] but did not increase the desire of participants to punish ($p=.26$),

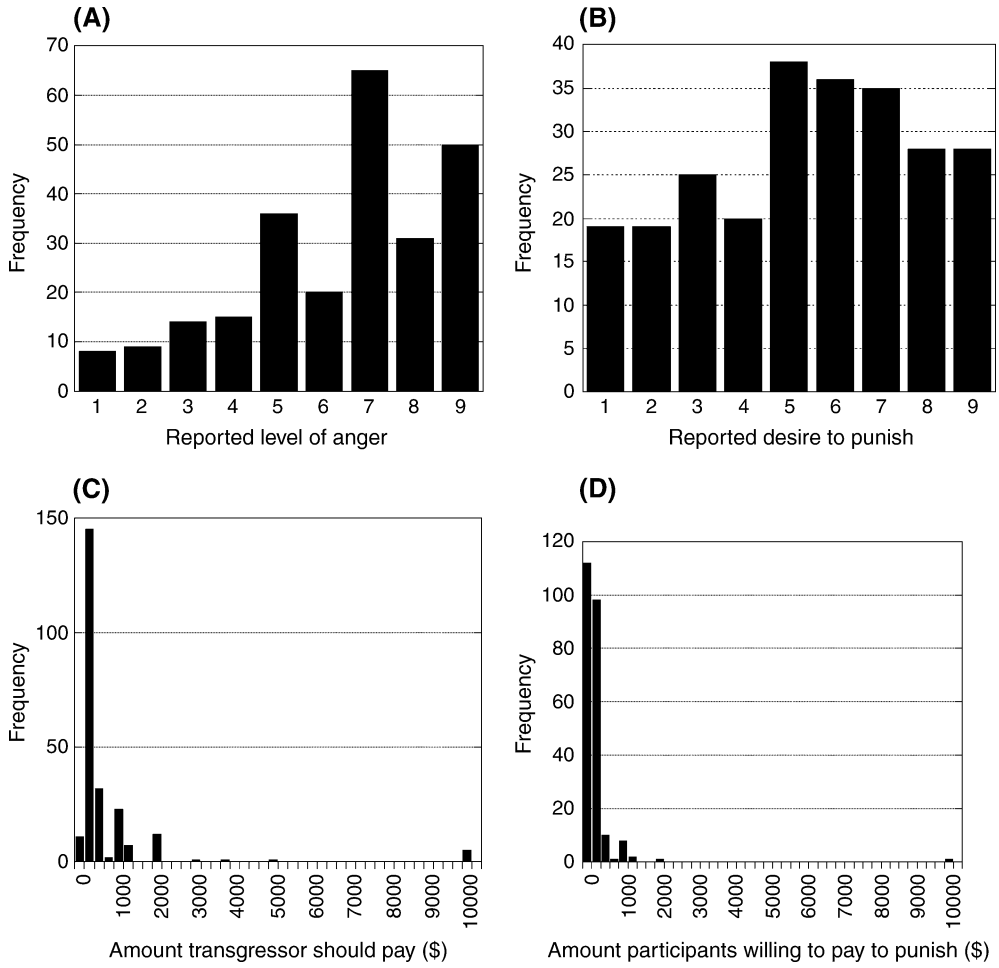

Fig. 2. Histograms of participants' responses to the fictional scenarios in Experiment 2 [n is 249, 249, 240, and 233 for (A), (B), (C), and (D), respectively]. Panel (A) shows the distribution of responses for reported levels of anger, showing that a majority of participants reported a high level of anger. Panel (B) shows that a substantial number of participants also had a high desire to punish the transgressor. Panel (C) shows the amount of money that participants thought the cheater should repay, with intervals of US$250 (the x axis labels show the upper limit of the corresponding interval). Panel (D) shows the amount of money that participants reported being willing to pay to punish the transgressor, with intervals of US$250 (the x axis labels show the upper limit of the corresponding interval). The data support the findings in Experiment 1. The less skewed data for Panel (A) contrast with the distribution for Experiment 1 [Panel (A)], which is due to the manipulation of the presence or absence of a normative option to punish the transgressor by going to small claims court.

the amount they thought the transgressor should pay (p=.83), or the amount that they were willing to pay to punish (p=.99).

There was a main effect for sex on desire to punish the transgressor [$F(2,238)$=5.74, p=.004; M_{Males}=6.0, S.D.$_{\text{Males}}$=2.4, M_{Females}=5.0, S.D.$_{\text{Females}}$=2.4] and near-significant effect on level of anger [$F(2,238)$=2.97, p=.05; M_{Males}=6.6, S.D.$_{\text{Males}}$=2.2, M_{Females}=6.2,

S.D.$_{Females}$=2.2]. This suggests that females are emotionally more moderate in their responses to the transgression, but this did not produce a sex effect for the amount that participants thought the transgressor should pay (p=.92) nor the amount that they were willing to pay to punish (p=.78). Furthermore, the stronger sex effect (on desire to punish the transgressor) did not produce notably different distributions for males and females for responses to that question.

In contrast to the punishment scenario, participants were highly sensitive to information regarding genetic relatedness and potential for future interactions in the helping scenario. Participants showed an overall moderately positive level of emotional sympathy for the individual needing help (M=6.7, S.D.=2.2), a moderate willingness to help (M=6.6, S.D.=2.2), and a mean offer of financial assistance of US$553.1 (S.D.=886.0). Table 1 presents a summary of participants' responses for each of the three dependent measures by perceived relatedness and potential for future interactions.

An analysis of variance showed that participants had greater sympathy for relatives versus strangers [F(1,125)=4.51, p=.036] and for protagonists with an expected future interaction versus no future interactions [F(1,125)=8.8, p=.004]. For willingness to help, participants made similar distinctions between relatives and strangers [F(1,125)=12.87, p<.0005] and between future and no-future situations [F(1,125)=12.37, p=.001]. There was also an interaction effect between those two variables [F(1,125)=9.37, p=.003], with the stranger/no-future interactions treatment significantly differing in post hoc analysis from the other three treatments. Finally, participants in the altruistic helping treatment made significant distinctions depending upon relatedness [F(1,117)=11.61, p=.001] for the amount that they were willing to provide as assistance (participants were willing to provide a mean of US$831.50 to a cousin but only US$305.10 to a stranger), but not depending upon the likelihood of future interactions [F(1,117)=0.18, p=.67]. There was a near-significant effect for sex on level of sympathy [F(2,118)=3.0, p=.054; M_{Males}=6.4, S.D.$_{Males}$=2.1, $M_{Females}$=7.2, S.D.$_{Females}$=2.0]. This suggests that females report feeling more sorry for the protagonist, but this did not produce a sex effect for how much participants thought they should help (p=.19) or how much they would contribute to help (p=.96).

Table 1
Mean values (S.D., n for cell) for responses to questions regarding the helping scenario

Dependent measure		Cousin	Stranger
Sympathy for protagonist	Likely future interactions	7.5 (1.7, 32)	7.1 (1.6, 31)
	Unlikely future interactions	6.8 (2.2, 33)	5.6 (2.6, 33)
Willingness to help	Likely future interactions	7.3 (1.8, 32)	7.1 (1.6, 31)
	Unlikely future interactions	7.1 (1.7, 33)	4.8 (2.6, 33)
Amount willing to give	Likely future interactions	768.97 (915.09, 29)	434.22 (291.3, 32)
	Unlikely future interactions	896.36 (1458.14, 28)	175.94 (184.06, 32)

The questions on sympathy for the protagonist and willingness to help the protagonist had a response scale of 1 to 9. The question asking how much help would participants provide to the protagonist, quantified as dollars, was open; that is, participants could suggest any amount that they wished. Note that while all participants were presented with the three questions, the cells present data for independent subjects, not repeated measures.

4. Discussion

Our fictional scenarios were designed to resemble a large body of research involving actual interactions and gave comparable results. In particular, some but not all participants indicated a willingness to punish at their own expense despite the absence of return benefits. Our experiments went beyond previous research by manipulating information regarding genetic relatedness, friendship, and potential for future interactions in a complete factorial design, which is more feasible for fictional scenarios than actual interactions. These are the most important variables influencing the evolution of cooperation and altruism according to kin selection theory (Hamilton, 1964, 1975) and evolutionary game theory (Axelrod & Hamilton, 1981; Maynard Smith, 1982). We therefore expected them to have a strong effect on the desire to altruistically punish and were surprised when they turned out to be statistically insignificant, despite large sample sizes. It is important to emphasize that the altruistic punishment scenario strongly engaged the interest of the participants, as indicated by their reported degree of anger, their desire to see the transgressor punished, and the willingness of at least some of the participants to punish at their own expense. The variables of genetic relatedness, friendship, and potential for future interactions had no effect despite a strong overall psychological response.

These results were so surprising to us that we decided to replicate the experiment with minor changes and to add an altruistic helping scenario. The second experiment confirmed the results of the first experiment with respect to altruistic punishment and demonstrated a strong sensitivity to information regarding genetic relatedness and potential for future interactions with respect to altruistic helping, as we originally expected for all forms of altruism. Thus, insensitivity to these major variables is not an artifact of our experimental procedure but indicates an important psychological difference between altruistic punishment and altruistic helping, at least in response to fictional scenarios.

We think that this is an important empirical result that demands an explanation, even if the correct explanation is not immediately apparent. We conclude with a brief outline of the major possibilities in hope of stimulating additional research.

First, our result might be an artifact of using fictional scenarios that have no relevance to real-world interactions. By itself, this possibility is not explanatory because it criticizes an entire method rather than our particular result does. Fictional scenarios are widely used in traditional and evolutionary psychological research and are closely related to questionnaire studies, which are even more widely used. If the method is to be used and trusted at all, it cannot be invoked to distrust a particular result. Moreover, it does not explain any of the details of our results, such as the difference between altruistic punishment and altruistic helping. It is clear that all research involving fictional scenarios and questionnaires is most instructive in conjunction with studies of real-world interactions and often reveal important psychological mechanisms, even when they do not correspond directly to behavior in actual interactions. In our case, our fictional scenarios are closely related to experiments involving actual interactions and give comparable results with respect to altruistic punishment, suggesting that the difference between altruistic punishment and altruistic helping is more than a meaningless artifact. See Wilson and O'Gorman (2003) for a more detailed discussion

of fictional scenarios as a research method and the importance of narrative in psychological and cultural processes.

Second, our result might reflect a mismatch between the modern environment and psychological mechanisms adapted to the ancestral environment, as suggested for altruistic punishment by Johnson et al. (2003). One version of the mismatch hypothesis maintains that all social interactions in the ancestral environment were among genetic relatives and nonrelatives with a high probability of future interactions. Because these were environmental constants, humans did not evolve the ability to discriminate and behave appropriately toward strangers with no possibility of future interactions. Not only is this claim implausible, based on what we know about the ancestral social environment (Fehr et al., 2002; Henrich, 2004), but it is undermined by our experimental results, which show that people are fully capable of discriminating information regarding genetic relatedness and potential for future interactions in the context of helping. Moreover, this particular version of the mismatch hypothesis cannot explain the pronounced individual differences observed in our experiments and previous research involving actual interactions. The mismatch hypothesis cannot be used in its general form to cast doubt on a specific result. A specific version of the mismatch hypothesis must be proposed that accounts for the experimental results in a testable fashion.

Third, our result might be explained in terms of adaptation and natural selection. Moral systems include some rules that are regarded as inviolable and others that are regarded as voluntary. The Ten Commandments provide examples of rules that are regarded as inviolable. The commandment "Thou shalt not steal" is not stated with a list of qualifiers such as "except from a stranger" or "except if you are about to leave the group." It has an unconditional quality similar to the response of our participants to the cheating scenario. Giving to charity provides an example of a voluntary rule, which is regarded as morally praiseworthy but is not required. It is understood that such voluntary acts can be made at the discretion of the actor, similar to the conditional response of our participants to the helping scenario.

Obviously, the so-called inviolable rules are broken in ways that can even be morally justified, such as to avoid a greater moral evil. Nevertheless, the distinction is important, especially at the psychological level, and probably has functional significance as well. Groups whose members conceptualize at least some rules as unconditional and therefore punishable when broken are likely to function better than other groups. To proceed further, it is necessary to identify the suite of psychological traits that is required for the conceptualization and enforcement of such rules and to see if they can plausibly evolve by natural selection. Theoretical models devoted to this subject are more recent than the earlier theories of kin selection and reciprocal altruism (although the original paper of Trivers, 1971, on reciprocal altruism includes relevant verbal discussion), but they are just as important for understanding human evolution (Boyd et al., 2003; Boyd & Richerson, 1992; Gintis, 2000; Henrich & Boyd, 2001; Sober & Wilson, 1998; Wilson, 2002).

Empirically, Boehm (1993, 1999) has argued that modern hunter–gatherer societies are first and foremost moral communities, with a strong sense of right and wrong that applies to everyone, regardless of their status. This normative social environment enforced by punishment creates a degree of behavioral uniformity within groups and differences among

groups (when they abide by different norms) that is highly favorable for the genetic and cultural evolution of within-group cooperation. It is precisely the traits associated with the conceptualization and enforcement of moral rules that expand the circle of cooperation beyond genetic relatives and narrow reciprocators, causing human evolution to embark upon a different trajectory than all other primate species.

To summarize, altruism can take place in the context of punishment in addition to helping behaviors. Altruistic punishment has been empirically demonstrated in humans in experiments involving actual interactions. Our experiments reveal a remarkable insensitivity to information regarding genetic relatedness, friendship, and potential for future interactions in the impulse to altruistically punish, in contrast to the impulse to altruistically help, at least in terms of the psychological response to fictional scenarios. This result is difficult to explain in terms of kin selection and narrow reciprocal altruism, but it is plausible in terms of the psychological traits associated with moral systems. Any human trait revealed by an experiment can potentially be explained as a methodological artifact, as a nonadaptive byproduct, or as an adaptation. We think that the phenomenon of altruistic punishment and its unconditional nature is most likely to be explained as part of a suite of psychological traits associated with moral systems that is very much a product of natural selection.

Acknowledgments

We thank R. Boyd, S. Bowles, A.B. Clark, E. Fehr, D. Fessler, H. Gintis, P. Richerson, and the EEB group at Binghamton University for helpful discussion.

Appendix A.1. The following is one version of the questionnaire used in Experiment 1

This survey is part of a study on human social behavior. It will require approximately five minutes of your time, which is greatly appreciated. Your participation is strictly voluntary. Please read through Part I and then turn the page over to complete Part II.

Part I. Try to imagine yourself as part of the following story and then answer the questions on the back of this page:

Suppose that you and nine friends [cousins] in your town decide to pool your money to invest in the stock market. Pooling the money allows the group to qualify for lower transaction costs than if each person invests separately. [*Alternative for stranger condition*: Suppose that you decide to join an investing club in your town whose members pool their money to invest in the stock market. Pooling the money allows the group to qualify for lower transaction costs than if each person invests separately. Ten people join the club, none of whom knew each other previously.] Each person agrees to contribute US$1000 to the pool and to equally share the profits or losses.

The stocks do very well and triple in value after only one year. Just before you meet to divide the profits, you discover that the person who volunteered to keep the books only

invested US$200, changing the records so that others would not notice. You do some calculations and determine the following facts:

1. The total amount invested was US$9200 which tripled in value to US$27,600.
2. This was divided equally among all 10 friends [cousins/individuals] to yield US$2760 for each person.
3. The person who contributed US$200 should have received only US$600 and therefore received US$2160 more than deserved.
4. Everyone else should have received US$3000, or US$240 more than they actually got.

2. One of the following paragraphs followed the preceding text, manipulating perception of likely future interactions

Condition 1: All of you live in the same town and this person is likely to associate with you and *your other friends* [your cousins/the other members of the club] in the future. How are you going to act?

Condition 2: You are moving to another town but this person is likely to associate with *your other friends* [your cousins/the other members of the club] in the future. How are you going to act?

Condition 3: This person is moving to another town and is unlikely to associate with you or *your friends* [your cousins/the other members of the club] in the future. How are you going to act?

References

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Boehm, C. (1993). Egalitarian society and reverse dominance hierarchy. *Current Anthropology*, *34*, 227–254.

Boehm, C. (1999). *Hierarchy in the forest*. Cambridge, MA: Harvard University Press.

Bowles, S., & Gintis, H. (2002). Homo reciprocans. *Nature*, *415*, 125–128.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 3531–3535.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*, 171–195.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*, 1–25.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*, 980–994.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*, 169–179.

Hamilton, W. D. (1964). Genetical evolution of social behavior. *Journal of Theoretical Biology*, *7*, 1–52.

Hamilton, W. D. (1975). Innate social aptitudes in man, an approach from evolutionary genetics. *Biosocial anthropology*. R. Fox, London: Malaby Press.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization*, *53*, 3–35.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, *208*, 79–89.

Johnson, D. D. P., Stopka, P., & Knights, S. (2003). The puzzle of human cooperation. *Nature*, *421*, 911–912.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.

Ostrom, E. R., Gardner, R., & Walker, J. M. (1994). *Rules, games, and common-pool resources*. Ann Arbor, MI: University of Michigan Press.

Sober, E., & Wilson, D. S. (1998). *Unto others: the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Wilson, D. S. (2002). *Darwin's Cathedral: evolution, religion, and the nature of society*. Chicago, IL: University of Chicago Press.

Wilson, D. S., & O'Gorman, R. (2003). Emotions and actions associated with norm-breaking events. *Human Nature*, *14*, 277–304.