

**Costly Punishment in the Ultimatum Game Evokes Moral Concern, in Particular when Framed
as Payoff Reduction**

Kimmo Eriksson^{1,2,†}, Pontus Strimling^{2,3,4}, Per A. Andersson^{2,3,5}, and Torun Lindholm⁶

¹Mälardalen University, School of Education, Culture and Communication, Västerås, Sweden

²Stockholm University, Centre for the Study of Cultural Evolution, Stockholm, Sweden

³Linköping University, Institute for Analytical Sociology, Norrköping, Sweden

⁴Institute for Futures Studies, Stockholm, Sweden

⁵Department for Management and Engineering Division of Economics, Linköping, Sweden

⁶Stockholm University, Department of Psychology, Stockholm, Sweden

[†]Corresponding author. Mälardalen University, School of Education, Culture and Communication,
Box 883, SE-72123 Västerås, Sweden. Email: kimmo.eriksson@mdh.se. Phone: +46 21 101533.

This is a postprint version of: Eriksson, K., Strimling, P., Andersson, P. A., & Lindholm, T. (2017).

Costly punishment in the ultimatum game evokes moral concern, in particular when framed as payoff
reduction. *Journal of Experimental Social Psychology*, 69, 59- 64.

<https://doi.org/10.1016/j.jesp.2016.09.004>

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Abstract

The ultimatum game is a common economic experiment in which some participants reject another's unfair offer of how to split some money, even though it leaves them both worse off. This costly behavior can be seen as enforcement of a fairness norm and has been labeled "altruistic punishment", suggesting that it is a moral thing to do. But is this behavior viewed as moral by participants? Is it viewed as punishment? And are the payoff consequences of the behavior sufficient to determine the answers to these questions? To investigate this we framed costly punishment in two different ways: either as rejection of an offer (the standard ultimatum game framing) or as reduction of payoff. In a series of paid and hypothetical experiments we found that moral concerns about costly punishment depended on the framing. Specifically, the reduction frame elicited more moral concern about, and less use of, costly punishment than did the rejection frame. Several implications are discussed.

Keywords. costly punishment; ultimatum game; norm enforcement; motives; framing; moral judgment

Costly Punishment in the Ultimatum Game Evokes Moral Concern, in Particular when Framed as Payoff Reduction

When someone does not respect norms about fairness, a peer may sanction the norm violator. Game theory has become a popular paradigm for behavioral and evolutionary studies of how such social control can make groups do better (e.g., Fehr & Gächter, 2000). A core idea in this paradigm is that, on the one hand, groups tend to benefit from fairness norms being enforced but, on the other hand, it is costly for the individual who enforces the norm. In game theoretic terms, norm enforcement is modeled as a strategic choice that has negative payoff for oneself as well as for the target but potential benefits for the group by promoting adherence to the social norm. Such choices are referred to as *costly punishment* in this literature. It is possible to interpret an act of costly punishment as an altruistic act, as there is a cost to the actor and a potential benefit to the group. This interpretation is the basis of the strong reciprocity hypothesis in evolutionary social science (e.g., Boyd et al., 2003; Fehr, Fischbacher, & Gächter, 2002; Fehr & Gächter, 2002). However, this is a controversial interpretation (e.g., Kurzban & DeScioli, 2013), which merits further investigation.

The literature on altruistic punishment relies to a large extent on experiments using two economic games. The *public goods game with punishment* (PGP) is a symmetric game in two stages, played in a group of players (Fehr & Gächter, 2002). The first stage is the public goods game, in which each player decides their own monetary contribution to a common project; contributions benefit the group but are costly for the individual. Then follows the punishment stage, in which players can reduce the payoff of selected other players at a cost to themselves. Usually the punishment-to-cost ratio is 3:1 (e.g., Herrmann et al., 2008). In contrast to the PGP, the *ultimatum game* (UG) is an asymmetric two-player game (Güth et al., 1982). The first player makes an offer to the second player about how to divide a sum of money. The second player has a choice between accepting this offer or rejecting it, in which case neither player gets the money.

Despite the differences between the two games, the term altruistic punishment has been used to refer both to costly reduction of the payoff of selfish players in the PGP (e.g., Fehr & Gächter, 2002; Herrmann et al., 2008) and to rejection of unfair offers in the UG (e.g., Crockett et al., 2010; Fehr &

Fischbacher, 2003; Strobel et al., 2011). This usage suggests that the acts of reduction and rejection are equivalent, in particular in terms of being altruistic and punitive. But are they?

“Rejection” vs. “Reduction” Framing of Costly Punishment of an Unfair Division

A substantial body of research demonstrates that behavior in economic games tends to be sensitive to how the game is framed (e.g., Brewer & Kramer, 1986; Eriksson & Strimling, 2014; Liberman, Samuels, & Ross, 2004). However, framing effects on the use of costly punishment have been largely ignored. To examine framing effects we shall use the ultimatum game and assume that the second player faces an offer of an unfair 3:1 division of, say, 100 dollars. Rejection of the offer then costs the first player three times more than the second player. Instead of framing the second player’s decision as whether to reject the unfair division we can therefore alternatively frame it as a decision about whether to pay 25 dollars to reduce the first player’s payoff by three times as much. These are the framings used in the current paper. They will be referred to as *Rejection* and *Reduction*, respectively.

In the experimental literature, UG experiments generally use a Rejection frame and PGP experiments use a Reduction frame. PGP experiments were explicitly designed to study costly punishment, whereas the UG was originally designed to study bargaining. The interpretation of rejection as punishment was not even mentioned in the classic paper introducing the UG (Güth et al., 1982). Generalizing from researchers to participants, could it be that punishment is less readily associated with the Rejection frame than with the Reduction frame? We shall return to this possibility in our Hypotheses section below.

Rejection vs. over-claiming. No prior work seems to have examined effects of framing costly punishment as Rejection or Reduction. However, an important study of framing effects in the UG compared the standard bargaining framing with a social dilemma framing (Larrick & Blount, 1997). In the social dilemma framing, the first player claimed some proportion of a resource to be divided; the second player, knowing the first player’s claim, also claimed a proportion, and the claims were paid out only if their sum did not exceed the entire resource. As an economic game this is equivalent to the UG, as the first player’s claim is equivalent to an offer of a division of the resource and an over-claim by the second player is equivalent to rejection of the offer. Larrick and Blount found that framing

mattered, such that the social dilemma frame elicited more cooperative behavior from both players than the bargaining frame. Moreover, over-claiming in the social dilemma frame was found to be judged as more selfish, less rational, and more vindictive than rejection in the bargaining frame.

Over-claiming is different from both rejection and reduction. According to Larrick and Blount (1997, p. 16), what makes the social dilemma frame special is that the second player's decision to claim an amount "is a relatively independent action" and "provides only an indirect means of expressing approval or disapproval". In contrast, both Rejection and Reduction are explicit reactions to the first player's behavior. Hence, although the study of Larrick and Blount demonstrates that the UG is sensitive to framing, it cannot tell us what would be the effect of framing costly punishment as Reduction instead of Rejection.

Hypotheses

The reason to expect framing effects. Above we hinted at a possible key difference between the Rejection and Reduction frames: Rejection may be less likely to be perceived as an act of punishment, or as directly causing harm to the other player. After all, people have considerable everyday experience of receiving various offers, say, of buying certain goods at a certain price. In general, decisions whether or not to accept such offers would presumably be guided by considerations of whether the offer is satisfactory rather than by considerations of whether the proposer should be punished. The fact that the proposer is worse off if the offer is rejected is likely to be seen as an indirect effect of rejection and not as its purpose. Consistent with this reasoning, Yamagishi et al. (2009) found that rejections of unfair splits were still fairly common even when second players could only reject their own share but not affect the first player's share.

The domain of framing effects: moral judgments, motives, and use of costly punishment. If costly punishment of unfair behavior in economic games is generally seen as "altruistic" it should be morally right to use it. However, costly punishment is by definition harmful. Punishment should therefore be morally right only if the harm to the norm violator is justified by the good inherent in the justice of a norm violator being punished. There are two reasons why would-be punishers in the UG may not see it as justified to harm the other. First, there is the issue of causing direct harm to another; research on other moral dilemmas shows that people tend to experience a strong negative emotional

reaction to the thought of causing direct harm, even to achieve a greater good, and therefore judge such actions as immoral (Greene et al., 2001). A second issue is that the good in this case is justice handed out by a peer, and people may have internalized norms against taking justice in their own hands, preferring to leave it to authorities.

We know of no data that speak directly to the moral status of costly punishment in the UG. In the PGP there is a growing body of work studying reactions to players who use costly punishment (e.g., Cinyabuguma, Page & Putterman, 2006; Kiyonari & Barclay, 2008; Nikiforakis, 2008; Strimling & Eriksson, 2014). Consistent with moral concerns about Reduction, these reactions have been found to be more negative than positive. However, by focusing on others' reactions to costly punishment this work has left unexplored whether players' moral concerns may affect their own motivation to use costly punishment and ultimately their actual behavior. We shall now discuss why we expect framing effects on moral concerns about costly punishment in the UG, as well as on motives for using and not using it, and on its actual use.

Predictions about moral judgments. Consistent with the abovementioned findings on Reduction in the PGP, we expect moral judgments of Reduction to be more negative than positive. We expect moral judgments of Rejection to be less negative, because the issues of causing direct harm and handing out justice should be less salient in the Rejection frame.

Predictions about motives. Participants in economic game experiments may display a diverse set of motives (Butler, Burbank, & Chisholm, 2011). Different motives may lie behind the exact same behavior (Hein et al., 2016). Prior research suggests that use of costly punishment can be motivated by a desire to punish as well by a desire to reduce inequality (Dawes et al., 2007). Based on our above analysis of the key difference between Rejection and Reduction, we predict the punitive motive to be more pronounced for Reduction.

We are not aware of any prior research on motives for *not* using costly punishment in economic games; to avoid the cost of punishment has simply been seen as the economically rational choice. By extension of our hypothesis about moral concerns we predict that there should also be a moral motive for refraining from costly punishment, in particular in the Reduction frame.

Predictions about behavior. According to the Logic of Appropriateness model for decision making in social dilemmas, behavior is evoked as a response to the question: “What does a person like me do in a situation like this?” (Weber, Kopelman, & Messick, 2004). In particular, a situation that creates moral concerns about one possible behavior should influence people who see themselves as moral to use another behavior. Since we predict greater moral concern about Reduction than about Rejection, it then follows that use of costly punishment should be less likely in the Reduction frame than in the Rejection frame.

Outline of Studies

We report a series of experiments conducted online with US based users of the Amazon Mechanical Turk. Informed consent was always obtained prior to participation.¹ The uniqueturker.myleott.com service was used to ensure that no subject took part in more than one of these studies. We report all measures, manipulations, and exclusions in these studies. Raw data are available as supplementary material.

Experiments 1 and 2

Our first couple of experiments were designed to examine the effect of framing costly punishment as Rejection vs. Reduction upon the choices people make and the motives they report. Experiment 1 used hypothetical decisions, which is cheap, so that we could afford a rather large sample size ($N = 400$) in order to achieve good estimates of effect sizes. As we shall see, a substantial framing effect was found. Experiment 2 replicated Experiment 1 in an actual game, hosted online, in which participants could gain up to \$4. In order to avoid excessive cost we used a smaller sample size ($N = 168$), with the aim of replicating the main findings of Experiment 1. The two studies used the same between-subject design (Reduction vs. Rejection) and the same measures. We therefore report them together.

¹ No ethical approval was sought, as the Swedish Act concerning the Ethical Review of Research Involving Humans (2003:460) states that studies with adults using informed consent need approval only if they use a method intended to physically or mentally influence a person or if they involve sensitive information that can be traced back to individual persons.

Method for Experiment 1

We recruited 400 US based users of the Amazon Mechanical Turk for a fee of 1.25 US dollar (50% females; ages ranged from 18 to 72 years, $M = 35$ years, $SD = 11$ years).

Participants were assigned to one of two conditions (Rejection or Reduction). In both conditions they were asked to imagine that they were taking part in an on-line experiment in which another anonymous player was asked to make an offer about how to split 100 dollars between them, without any work required—and had decided on an offer of 25 dollars to the participant. In the Rejection condition, participants were told they could accept or reject this offer. In the Reduction condition, participants were told they could pay the 25 dollars they had earned in the experiment to reduce the other player's payoff by three times as much.

After a question about what they would feel towards the other player in this situation (not analyzed here) participants were asked what they would do, using a binary response set (i.e., reject/accept or reduce/not reduce depending on condition). Participants were then asked for their primary motive for the choice they just made. The response set, assembled based on theoretical considerations and previous work on the ultimatum game by Yamagishi et al. (2009, 2012), was: *to punish the proposer, to be fair, it's the morally right thing to do, to earn as much money as possible, to defy the proposer, to protect myself, other reason, don't know*. The same motives were available regardless of what choice had been made.

The experiment continued with various other measures not analyzed here. They are listed in the Supplementary Material together with the exact instructions for this experiment.

Method for Experiment 2

We recruited 168 US based users of the Amazon Mechanical Turk (50% females; ages ranged from 18 to 58 years, $M = 32$ years, $SD = 10$ years). Potential participants were instructed that they would take part in an online experiment where they would be paired with another participant. The reward for participation was stated as a guaranteed 2 US dollars, with a chance to earn another few dollars depending on decisions made.

At the experiment website participants were required to log in via their Facebook accounts, as a means to create unique user accounts. Because this login procedure might influence participants' sense

of anonymity, they were clearly informed that their login data would not be used in the study and promised that they would remain completely anonymous. Participants were put in an online waiting room until another participant joined, at which time the experiment began.

We were only interested in second-player decisions, and only in response to 3:1 splits from the first player. To make the game a proper two-player game without sacrificing half the participants, every participant made decisions in both roles. First came the second-player decision in response to a 3:1 split from another player. In the subsequent first-player decision, participants were constrained to the single option of offering a 3:1 split. Importantly, this constraint on the part of first players was not known to participants when they made their second-player decision.

As participants started the experiment they were randomly assigned to a condition, either Rejection or Reduction. In both conditions they were told they had been assigned “player type B”, meaning that the other player would make the decision of distributing 2 US dollars between them. They were then informed that the other player had made the decision of keeping 1.5 dollars and offering/giving 0.50 dollar. Participants were first asked to rate their feelings after receiving this decision (not analyzed here). Participants in the Reduction condition were then given the opportunity to reduce the other player’s payoff by 1.5 dollars at a cost of 0.5 dollars to themselves. Participants in the Rejection condition were instead given the opportunity to reject the proposal. In both conditions participants were asked to motivate their decision, with the same response options as in Experiment 1.

Participants were then informed that the roles would now be switched and another round of the game would be played. They were instructed that they were unfortunately not given any choice but to keep 1.5 dollars for themselves and to offer 0.50 dollar, as this was the distribution of interest in the study. The other player’s response to that pre-set decision was then presented. Total earnings were calculated and displayed on a final result screen. All participants then completed a battery of personality questionnaires that are not analyzed here (see list in Supplementary material).

Results for Experiments 1 and 2

In Experiment 1 a few participants did not complete all measures, hence slightly lower sample sizes than 200 per condition are reported in the statistical tests below.

Frequency of use of Rejection/Reduction. As reported in Table 1, use of costly punishment was much less common when framed as Reduction than when framed as Rejection (Exp. 1: $\chi^2(1, N = 392) = 20.89, p < .001, OR = 2.82$; Exp. 2: $\chi^2(1, N = 168) = 4.82, p = .028, OR = 2.11$).

Table 1
Frequencies of Use of Rejection/Reduction in Experiments 1-2

Experiment	Rejection	Reduction
1	41% (n = 195)	20% (n = 197)
2	39% (n = 66)	24% (n = 102)

Note. Reported n-values refer to the number of participants who completed the decision of whether to use Rejection/Reduction in each experiment. Percentages refer to the proportion of these participants who decided to use Rejection/Reduction.

The moral motive. Table 2 reports how often the motive “it’s the morally right thing to do” was given for use and non-use of Rejection and Reduction. The table shows a clear and consistent pattern. First, the moral motive was seldom given among those who *used* costly punishment, whether framed as Rejection or Reduction. Second, the moral motive was frequent among those who *refrained* from using Reduction but not among those who refrained from using Rejection. The association between frame and a moral motive for refraining from costly punishment was highly significant in Experiment 1, $\chi^2(1, N = 273) = 35.12, p < .001$, as well as in Experiment 2, $\chi^2(1, N = 118) = 15.45, p < .001$. Odds ratios could not be calculated as certain frequencies were zero (i.e., the odds ratios were infinite).

Table 2
Frequencies of a Moral Motive for Using and Not Using Rejection or Reduction

	Experiment 1		Experiment 2	
	Rejection (total n = 195)	Reduction (total n = 197)	Rejection (total n = 66)	Reduction (total n = 102)
Among users	8% (n = 80)	0% (n = 39)	8% (n = 26)	0% (n = 24)
Among non-users	0% (n = 115)	26% (n = 158)	0% (n = 40)	31% (n = 78)

Note. Reported n-values refer to the number of participants who decided to use (top row) or not use (bottom row) Rejection/Reduction in each experiment. Percentages refer to the proportion of these participants who motivated their decision as the moral thing to do.

Punishment and fairness as motives for using Rejection/Reduction. In both experiments, and regardless of whether costly punishment was framed as Rejection or Reduction, the two most common motives for using it was “to punish” and “to be fair”. However, as shown in Table 3, which of these two motives was most common differed between conditions. Whereas “to punish” was the more common motive for Reduction, “to be fair” was the more common motive for Rejection. The association between frame and motive was significant for Experiment 1, $\chi^2(1, N = 72) = 6.61, p = .010, OR = 3.67$, and marginally significant for the less powered Experiment 2, $\chi^2(1, N = 41) = 3.35, p = .067, OR = 3.36$. Together the two studies provide good evidence for punitive motives being more strongly elicited in the Reduction frame.

A complete summary of the data on motives are given in Tables S1-2 in the Supplementary Material.

Table 3

Frequencies of Reported Motives among Those Who Used Rejection/Reduction

Motive	Experiment 1		Experiment 2	
	Rejection (n = 80)	Reduction (n = 39)	Rejection (n = 26)	Reduction (n = 24)
To punish the other	22%	59%	38%	58%
To be fair	29%	21%	46%	21%

Note. Reported n-values refer to the number of participants who decided to use Rejection/Reduction in each experiment.

Discussion

Across paid and hypothetical decisions we found that Reduction was used much less frequently than Rejection; non-use of Reduction was more often morally motivated; and use of Reduction was more often punitively motivated. These effects were all in the direction we predicted.

Several limitations should be noted. For instance, participants motivated the choice they made but we do not know what motives it was weighed against. Moreover, the response set on motives did

not include all theoretically relevant motives for *not* using costly punishment (in particular, “It is not my role to mete out punishment” was missing). A third limitation is that participants could only choose one motive but possibly their motives were more complex. A final limitation is that the framing manipulation differed not only in whether the act was described as rejection or reduction but also in the description of the payoff consequences as either absence of gain or loss – “not receive” (Rejection) vs. “lose” (Reduction). Thus the framing effect could be due to loss aversion rather than the mechanisms outlined in our hypothesis.

Experiment 3

We conducted a third experiment to address the limitations listed above. Different response sets were used for motives for using and motives for not using costly punishment. Participants were allowed to express more complex motives by choosing a secondary motive to complement their primary choice, and they were also asked for the most relevant motives for making the choice they did *not* make in the end. Finally, the description of payoff consequences of Rejection was changed to say that both participants then “lose the money”. We used hypothetical decisions as the first couple of experiments indicated that results were similar for paid and hypothetical decisions.

Method

We recruited 200 US participants on the Amazon Mechanical Turk (46% females; mean age 34 years, $SD = 12$). 100 participants were assigned to each of two conditions. The conditions were similar to Experiment 1 except for the changes outlined above. See Supplementary Material for full instructions.

Results and Discussion

198 out of 200 participants made a choice about whether they would use Rejection/Reduction. Rejection (32%) was more common than Reduction (20%), $\chi^2(1, N = 198) = 3.76, p = .053, OR = 1.89$. Although only marginally significant in this experiment, the effect was quite consistent with the results of the previous experiments that were reported in Table 1.

A complete summary of the data on motives are given in Table S3 in the Supplementary Material. Here we focus on how often the theoretically most relevant motives were selected as the primary motive for using or not using Rejection/Reduction.

Moral motives. Results on moral motives were consistent with the previous experiments and our hypothesis. Thus, it was uncommon to give the primary motive “it seems morally right” for using Rejection (3%) or Reduction (4%). The primary motive that “it seems morally wrong” to use Rejection/Reduction was more common for refraining from Reduction (19%) than for refraining from Rejection (4%), $\chi^2(1, N = 200) = 11.05, p = .001, OR = 5.63$. In addition, the primary motive “It is not my role to mete out punishment” was also more common for refraining from Reduction (14%) than from Rejection (5%), $\chi^2(1, N = 200) = 4.71, p = .030, OR = 3.09$.

Punishment and fairness as motives for using Rejection/Reduction. The primary motive “to punish” was more common for using Reduction (43%) than for using Rejection (24%), $\chi^2(1, N = 200) = 8.10, p = .004, OR = 2.39$. Consistent with the previous finding, the primary motive “to achieve a fair outcome” was less common for using Reduction (24%) than for using Rejection (51%), $\chi^2(1, N = 200) = 15.55, p < .001, OR = 3.30$.

Experiments 4 and 5

The results on motives in the first three experiments are consistent with the hypothesis that Rejection is less readily interpreted as punishment and that Reduction elicits stronger moral concerns. However, many participants chose other motives altogether. To test the hypotheses about interpretations and moral concerns more directly we conducted two more data collections, which we report together.

Method

For each experiment, 200 unique US participants were recruited on the Amazon Mechanical Turk (in total 46% females, mean age 34 years, $SD = 10$). In each experiment, 100 participants were assigned to each condition (Rejection or Reduction). Similar to the previous experiments, participants were presented with a description of either a Rejection or a Reduction scenario. See Supplementary Materials for exact instructions.

In Experiment 4, participants were asked how they would interpret someone using Rejection/Reduction. They were given a choice between four response options: *The receiver is angry at the proposer and makes this decision to punish him/her* (coded as Punishment); *The receiver prefers a fair outcome to an unfair outcome* (coded as Fairness); *The receiver dislikes having the outcome decided by another and wants to assert his/her independence* (coded as Independence); *The receiver has another motive than any of the above* (coded as Other).

In Experiment 5, participants were asked what would be the morally right thing for them to do if they were the second player in the scenario. Responses were given on a five-point scale, where 1 represented that use of Rejection/Reduction is definitely *not* the morally right thing to do, and 5 represented that use of Rejection/Reduction is definitely the morally right thing to do.

Results

Interpretation of Rejection/Reduction. In support of our hypothesis, the punishment interpretation was much less common for Rejection (37%) than for Reduction (66%), $\chi^2(1, N = 200) = 16.83, p < .001, OR = 3.31$. A complete summary of the data on interpretations is given in Table S4 in the Supplementary Material.

Moral judgments. The median moral judgment of Rejection was at the scale mid-point of 3 (i.e., neutral), whereas the median moral judgment of Reduction was 2 (i.e., more critical). According to a Mann-Whitney U test the difference was statistically significant (mean rank 115.27 vs. 85.73), $U = 3.72, p < .001, r = .26$. A supplementary experiment using another measure of moral judgment gave the same result, see Supplementary Material.

General Discussion

In the introduction we pointed out that the notion of “altruistic punishment” is often used to refer to rejection of unfair offers in the ultimatum game. The results of our studies indicate that it is problematic to do so. For one thing, the association between rejection and altruism is doubtful, at least if altruism is perceived as a morally superior behavior. Specifically, we found little moral motivation for using rejection, and moral judgments of rejection were on average neutral. For another, the association between rejection and punishment is doubtful. Specifically, we found that rejection was typically not interpreted as punishment, and that the most common motive to use it was not to punish

but to achieve a fair outcome. Indeed, it is a fallacy to assume that because an act has a certain consequence, that consequence was the reason for the act (Kurzban & DeScioli, 2013). For this reason, we think researchers should be wary to use ultimatum game data as a basis for psychological theories of punishment (e.g., McCullough et al., 2013).

The term “altruistic punishment” is also used in reference to decisions of explicitly paying a cost to reduce the payoff of another player who has behaved selfishly. Our studies indicate that this operationalization is problematic too, but not in exactly the same way. Compared to rejection, reduction was more strongly associated with punishment – but at the same time its association with moral behavior was negative. Thus, whereas rejection tended to be viewed neither as altruistic nor as punishment, reduction tended to be viewed as immoral punishment. What makes it immoral, we believe, is that the moral pros of reducing the other player’s payoff (namely, that a violation of fairness does not go unchallenged) is outweighed by the moral cons (causing harm and disrupting social order).

Our approach contrasts with approaches based on (evolutionary) game theory, which assume that the strategic structure of a situation is the frame within which behavior should be understood. It has been argued that framing effects in games can be explained within the game theoretical paradigm as influencing players’ beliefs about how others will play (Dufwenberg et al., 2011). However, for decisions where payoffs do not depend on others, this argument is unable to account for framing effects. Such framing effects include those demonstrated in the present paper, as well as those found in earlier work on the second player in the ultimatum game (Larrick & Blount, 1997) and the dictator game (DeScioli, & Krishna, 2013). Our effects were obtained using two established framings that are considered as neutrally worded in this literature, suggesting that it is not worthwhile to strive for neutral language instead of purposefully inducing a particular frame (see also Eriksson & Strimling, 2014).

Cross-cultural studies of costly punishment in economic games demonstrate that there are substantial differences in behavioral patterns between cultures (Henrich et al., 2006; Herrmann, Thöni, & Gächter, 2008). One implication of our work is that cultural differences in the moral concerns evoked by the game context could be a factor behind differences in behavior.

Ultimately, the goal is to understand informal punishment in real life; Nisbett and Cohen's (1996) work on the culture of honor is a good example of such research. The norms about real-life informal punishment have shown some structural consistency across contexts and cultures (Strimling & Eriksson, 2014). For this reason we think it is likely that the qualitative difference between rejection and reduction holds also in contexts outside of economic experiments. Assuming it does, several interesting implications would follow. For one thing we would expect actions that are seen as rejection to be used more often, and be less disapproved of, than actions that are seen as reduction. For instance, withdrawing from a person you dislike would be more approved of than actively taking something from the person that he/she values. Avoidance and exclusion should be more common than direct punishment as ways to react to selfish behavior. Moreover, in real life actors can take an active part in shaping the framing. Assuming that people tend to feel moral concerns about actions that amount to reduction, they should make efforts to frame such actions as unintentional or beyond their control. To test these predictions it is of course necessary to study real-life behaviors instead of economic games.

Acknowledgments

This research was funded by the Swedish Research Council [grants 2008-2370, 2009-2390, 2009-2678, 2013-7681, 2013-5460], the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 324233, Riksbankens Jubileumsfond (M12-0301:1), and the Knut and Alice Wallenberg Foundation [grant number 2015.0005]. The funding bodies had no involvement in the research.

References

- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences, 100*, 3531-3535.
- Brewer, M. B., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology, 50*, 543-549.

- Butler, D. J., Burbank, V. K., & Chisholm, J. S. (2011). The frames behind the games: Player's perceptions of prisoners dilemma, chicken, dictator, and ultimatum games. *The Journal of Socio-Economics*, 40(2), 103-114.
- Cinyabuguma, M., Page, T., & Putterman L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, 9, 265-279. doi:10.1007/s10683-006-9127-z
- Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry*, 12(2), 128-143.
- Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G., & Robbins, T. W. (2010). Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion*, 10, 855-862.
- Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, 446(7137), 794-796.
- DeScioli, P., & Krishna, S. (2013). Giving to whom? Altruism in different types of relationships. *Journal of Economic Psychology*, 34, 218-228.
- Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73, 459-478.
- Eriksson, K., & Strimling, P. (2014). Spontaneous associations and label framing have similar effects in the public goods game. *Judgment and Decision Making*, 9, 360-372.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425, 785-791.
- Fehr, E. & Gächter, S. (2002) Altruistic punishment in humans. *Nature*, 415, 137-140.
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13, 1-25.
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3, 367-388.
- Hein, G., Morishima, Y., Leiberg, S., Sul, S., & Fehr, E. (2016). The brain's functional network architecture reveals human motives. *Science*, 351(6277), 1074-1078.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... & Ziker, J. (2006).

Costly punishment across human societies. *Science*, *312*, 1767-1770.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*, 1362-1367.

Kiyonari, T. and Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, *95*, 826-842. doi:10.1037/a0011381

Larrick, R. P., & Blount, S. (1997). The claiming effect: Why players are more generous in social dilemmas than in ultimatum games. *Journal of Personality and Social Psychology*, *72*(4), 810-825.

Lieberman, V., Samuels, S.M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining Prisoner's dilemma game moves. *Personality and Social Psychology Bulletin*, *30*, 1175–1185.

Kurzban, R., & DeScioli, P. (2013). Adaptationist punishment in humans. *Journal of Bioeconomics*, *15*, 269-279.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*, *36*, 1-15.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves?. *Journal of Public Economics*, *92*, 91-112.

[doi:10.1016/j.jpubeco.2007.04.008](https://doi.org/10.1016/j.jpubeco.2007.04.008)

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: the psychology of violence in the south*. Westview Press.

Strimling, P., & Eriksson, K. (2014). Regulating the regulation: Norms about how people may punish each other. In P. van Lange, T. Yamagishi, & B. Rockenbach (Eds.), *Reward and Punishment in Social Dilemmas* (pp. 52-69). Oxford University Press.

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *Neuroimage*, *54*(1), 671-680.

Weber, J. M., Kopelman, S., & Messick, D. M. (2004). A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review*, 8(3), 281-307.

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., & Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proceedings of the National Academy of Sciences*, 106, 11520-11523.

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., ... & Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences*, 109, 20364-20368.