

### Original Article

## The Effects of Perceived Anonymity on Altruistic Punishment

Jared Piazza, Institute of Cognition and Culture, Queen's University, Belfast, UK. Email: [jpiazza02@qub.ac.uk](mailto:jpiazza02@qub.ac.uk) (corresponding author)

Jesse M. Bering, Institute of Cognition and Culture, Queen's University, Belfast, UK

**Abstract:** Previous studies investigating altruistic punishment have confounded the effects of two independent variables: information transmission (or breach of privacy) and personal identification (or breach of anonymity). Here we report findings from a brief study in which participants were asked to respond to a social norm violation (i.e., an anonymous actor had behaved selfishly in an economic game) by deciding whether to sacrifice their own endowment to punish this person. A third of the participants were told that their economic decisions would be made known to another player but could not be identified (privacy breach condition), whereas another third were informed that their decision as well as their names would be made known (anonymity breach condition). (The decisions of control participants were completely anonymous and private.) Participants also justified their economic decisions and reported their emotional experiences. The results were participants punished most in the privacy and anonymity breach conditions and least in the control condition. These findings have implications for existing evolutionary accounts of altruistic punishment.

**Keywords:** Altruistic punishment, anonymity, privacy, reputation, third-party punishment game

---

### Introduction

Past research has shown that anonymity encourages selfish and anti-normative behavior because anonymity decreases the “social distance” between actor and recipient (Bohnet and Frey, 1999; Burnham, 2003; Hoffman, McCabe, and Smith, 1996). However, much of this research has focused on adherence to norms of fairness, for example, norms about distributing endowed sums of money (Eckel and Grossman, 1996a; Hoffman et al., 1996). In comparison, less work has examined the effects of anonymity on punishment behavior. In other words, to what extent are people's decisions to punish others influenced by their concerns about others being able to identify them?

One study that did address this question of anonymity and punishment was conducted by Kurzban, DeScioli, and O'Brien (2007). In this study, these researchers informed participants, prior to making their decisions about whether to punish another player who has defected in a Prisoner's Dilemma Game, that either: (a) they would have to announce their decisions in public at the end of the game, (b) the experimenter would know their decisions, or (c) their decisions would be kept completely anonymous and confidential. Kurzban et al. observed increased levels of altruistic punishment (in the form of participants sacrificing more of their own monetary endowment to penalize non-cooperative players) when participants expected their decisions to be made public and their identities revealed.

Kurzban et al.'s (2007) results suggest that earlier findings about the role of anonymity in adherence to a fairness norm apply also to norms about punishing violators of those norms. However, their critical experimental treatment (of informing participants that their economic decisions would be made public, along with their identity) confounded two separate independent variables: (1) personal identification (or a breach of anonymity), and (2) information transmission (or a breach of privacy). While a breach of anonymity entails a breach of privacy, a breach of privacy does not necessarily entail a breach of anonymity. In the present study, we examined separately the effects of a breach of anonymity and a breach of privacy on altruistic punishment behavior in the context of a modified third-party punishment game.

Our main prediction in the present study was that participants would respond to a breach of privacy (e.g., being told that their names would be kept confidential but that their decision to punish would be revealed to another player) as though it were also breach of anonymity (e.g., that they would be personally identified as making this decision). This prediction is based on the more general evolutionary hypothesis that, in the ancestral past, once information about a social act was revealed and began to circulate through gossip, the actor's identity would have likely also been compromised in small-scale hunter-gatherer societies.

#### *Privacy, anonymity and social behavior*

Privacy is not the same as anonymity. Although privacy generally entails anonymity, anonymity does not necessarily entail privacy (although it is an open question whether people treat it as such). If someone finds an anonymous diary (i.e., private notebook) and shares its contents with others, this is a breach of privacy for the author of the diary, but it is not a breach of anonymity because the author is not identified. However, if the author of the diary can be identified due to the handwriting, this breach in privacy would now become also a breach of anonymity.

Several lines of research have shown that anonymity (in the sense of not being identifiable) often promotes antisocial and anti-normative behavior. For example, Chiou (2007) had Taiwanese youth respond to three sexual messages (interspersed with six non-sexual messages) over the computer, in a web camera (low anonymity), personal photo (medium anonymity), or nickname (high anonymity) condition. Chiou found that participants' responses to sexual messages were most explicit in the high anonymity condition, lowest in the low anonymity condition, and moderate in the medium anonymity condition. Likewise, Silke (2003) observed that public acts of violence by paramilitary foot-soldiers in Northern Ireland were exacerbated in terms of seriousness and

aggressiveness when perpetrators took measures to conceal their personal identities (wearing masks).

On the other hand, recent findings from evolutionary game theory have demonstrated the importance of social accountability (i.e., indirect reciprocity or reputation management) for the stabilizing of norm conformity and prosociality (Milinski, Semmann, and Krambeck, 2002; Nowak and Sigmund, 1998, 2005; Panchanathan and Boyd, 2004; Semmann, Krambeck, and Milinski, 2004; Wedekind and Braithwaite, 2002; Wedekind and Milinski, 2000). For example, Milinski et al. (2002) found that when information about participants' past decisions in an indirect reciprocity game were made known to other players, contributions to a public good were significantly higher than when participants' past decisions were unavailable. Moreover, when the opportunity to build a reputation is available and there are a limited number of potential interactive partners, people compete for the best reputation through costly (i.e., altruistic) giving (Barclay and Willer, 2007).

One possible explanation for why anonymity often loosens adherence to social norms is that cues that motivate reputation-enhancing behavior (i.e., cues indicative of being visible to others) are often absent in “anonymous” settings. Evidence for this hypothesis comes from several lines of research that have demonstrated increased prosocial behavior as a result of introducing implicit, social-presence cues (e.g., human-like eyes, human voice) into an ostensibly “anonymous” setting (Bateson, Nettle, and Roberts, 2006; Burnham and Hare, 2007; Haley and Fessler, 2005; Kurzban, 2001).

Burnham (2003), for example, devised a Dictator Game experiment with three conditions—a “no photo” treatment, a “dictator photo” treatment, and a “recipient photo” treatment. The Dictator Game assigns one player the role of “dictator” and another player the role of “recipient.” The dictator is given a sum of money and must decide how much of it he/she will share with the recipient. The recipient is simply a passive recipient in the game and has no sanctioning power over the dictator (cf. the Ultimatum Game). In Burnham's “no photo” treatment, participants simply performed an anonymous Dictator Game. In the “dictator photo” treatment, an instant photograph of the dictator was enclosed along with their decision to the recipient; that is, in this condition participants performing the role of dictator believed that they could be identified by recipients (which raised the possibility of interactions outside the laboratory). In the “recipient photo” treatment, dictators viewed an instant photograph of the recipient prior to their decision; that is, in this condition, participants performing the role of dictator could identify the recipient of their action (which introduced social presence cues into the “anonymous” environment). The outcome was that dictators were most selfish (i.e., they gave the least amount to the recipient) in the “no photo” (i.e., true anonymity) condition. However, in the two experimental conditions, there were significantly more equitable and generous offers, suggesting that both the presence of social cues (“recipient photo”) and concerns about being identified (“dictator photo”) promote generous behavior.

Thus, while researchers have investigated the effects of perceived anonymity on social behavior, it remains unclear the extent to which perceptions of privacy motivate similar behaviors. In particular, we are not aware of any study that has distinguished methodologically between a breach of privacy (i.e., knowing that others will learn about one's actions but that one's identity will be protected) and a breach of anonymity (i.e., knowing that others will learn about one's actions and also that one will be identified as the actor).

*Evolutionary accounts of altruistic punishment and the present study*

Norms of cooperation and fairness may evolve only when enough people are committed to punishing norm violations (Henrich and Boyd, 2001). However, punishment itself represents a second-order public good (Boyd, Gintis, Bowles, and Richerson, 2003; Yamagishi, 1986)—i.e., the benefits of punishment can be enjoyed by all members of a group regardless of individual contributions. Because the act of punishing norm violators is costly (i.e., one must expend time and energy to punish, and one risks retaliation and possible injury), it is in an individual's best interest not to punish, so as to avoid the costs of punishing. Nevertheless, people often do punish norm violators (at least in laboratory experiments), despite the costs and even when the actions of norm violators do not directly affect them (Fehr and Fischbacher, 2004; Fehr and Gächter, 2002; Henrich et al., 2006).

Gintis (2000) and Henrich and Boyd (2001) have argued that punitive dispositions evolved through group selection. Henrich and Boyd (2001), for example, presented a multi-stage model whereby once norms for cooperation and punishment stabilize within a “pay-off biased” and “weak conformist transmission” population, they may invade other populations and come to dominant by avoiding costly punishment (see also Boyd et al., 2003; Gintis, Bowles, Boyd, and Fehr, 2003). Another possibility is that individuals engaging in altruistic punishment were rewarded from third parties (i.e., group members) in the form of trust (Barclay, 2006), social status (Hardy and Van Vugt, 2006), or reproductive benefits (Miller, 2000; Smith, Bliege Bird, and Bird, 2003), which made it beneficial (in the long run) for individuals to punish norm violations.

Reputation-based explanations for altruistic punishment appear to fall short in that they fail to account for observations of altruistic punishment in anonymous settings, where there are no ostensible opportunities to build a reputation (e.g., Fehr and Gächter, 2002; Fehr and Fischbacher, 2004). This kind of altruistic behavior (that occurs between non-kin in the absence of cues for direct or indirect reciprocity) has been labeled “strong reciprocity” (Gintis, 2000; Fehr, Fischbacher, and Gächter, 2002) and there are principally two evolutionary accounts for this behavior. According to Gintis et al. (2003), strong reciprocity is an adaptive dispositional trait, which evolved through group selection. According to Burnham and Johnson (2005), strong reciprocity is, by definition, maladaptive, and they suggest that strong reciprocity (as it has been operationalized in laboratory experiments) may be caused by inadvertent “triggering” of individual-level indirect-reciprocity mechanisms under presumably anonymous conditions (referred to as the “evolutionary legacy hypothesis”; see Haley and Fessler, 2005, for a similar argument).

In the present study, all participants were assigned the role of the “third party” in a third-party punishment game (a one-shot economic game; see Fehr and Fischbacher, 2004) and had to decide how much to spend of their own endowment to punish an anonymous dictator who “selfishly” distributed £0 to an anonymous recipient. We manipulated participants' perceptions about having their punishment decisions revealed to the anonymous recipient (privacy breach) vs. having their decisions revealed and their identities revealed to the recipient (anonymity breach). Participants were randomly assigned to one of three experimental conditions. In the control condition, participants were given the impression that their economic decisions would be kept private and their identities kept strictly anonymous. In the privacy breach condition, participants were told that the anonymous recipient would hear about their economic decision, but would not

know their names (thus had no way of identifying them). Finally, in the anonymity breach condition, participants were informed that the recipient would be made aware of both their economic decision and their names (and thus could potentially identify them).

No significant difference in punishment behavior was expected between the privacy breach and anonymity breach conditions. However, altruistic punishment in these conditions was expected to be greater than in the control condition (i.e.,  $H_A$ : anonymity breach = confidentiality breach > control). This prediction follows from the hypothesis that concerns about reputation are inadvertently activated by the privacy breach cue (despite ostensible anonymity). Alternatively, participants may strategically reason that their economic behavior is equally anonymous in the privacy breach and control conditions, and thus adjust their behavior accordingly. If this is the case, there should be no difference in punishment behavior between the privacy breach and control conditions, with greater levels of punishment in the anonymity breach condition (i.e.,  $H_B$ : anonymity breach > privacy breach = control).

We also had participants justify their economic decisions in writing and report, by selecting from a list of emotion terms, those emotions they experienced while making their decision. In regards to participants' justifications, we expected that participants who punished the dictator would justify their actions in terms of concerns about fairness and justice, as suggested by previous findings (e.g., Dawes, Fowler, Johnson, McElreath et al., 2007; Hoffman et al., 1996). We made no predictions about justifications of non-punishers. Finally, in regards to participants' reports of their emotional experience, we expected punishers to report having had experienced the negative social emotions of anger/outrage, and the positive social emotions of pride/righteousness. From an appraisal perspective, anger/outrage arises from perceiving a negative event or social transgression as certain, under human control, and brought about by others (Lerner and Keltner, 2000; Smith and Ellsworth, 1985). Pride/righteousness is thought to arise from perceptions of the self as responsible for a positive event or adhering to a social norm (Fessler and Haley, 2003; Tracy and Robins, 2004). On the other hand, non-punishers were expected to report feeling the negative social emotions of guilt and shame. Guilt/shame is thought to arise from perceptions of the self as responsible for a negative (usually interpersonal) event or norm violation (Baumeister, Stillwell, and Heatherton, 1994; Tangney, Stuewig, and Mashek, 2007; Tracy and Robins, 2006).

## **Materials and Methods**

Forty-two students from Queen's University, Belfast, Northern Ireland (21 females, 21 males; 69% Caucasian Irish, 28.6% Caucasian English, 2.4% Irish/English; mean age = 20.81; min. = 18.42, max. = 36.25) were assigned randomly to one of three experimental conditions. In all three conditions, participants were instructed that they were the "third party" in a "three-player economic game." The economic game was said to entail one player (the dictator) who was to distribute a sum of ten £1 coins between him- or herself and a second player (the recipient). In other words, this was a Dictator Game in which the dictator had absolute control over how many coins (if any) the recipient would receive. The participants were instructed that these two other players were seated in separate rooms and that the dictator would be making his or her decision in private. In actuality, aside from the participant there were no other players. [Note that, in the study, the dictator was referred to

as “distributor” to avoid connotations associated with “dictator” (e.g., “cruel,” “tyrant”). Throughout this paper, however, we use the conventional moniker of “dictator.”]

As third party, participants were given a total sum of five £1 coins and informed that every £1 coin they “left behind” (i.e., that they did not take with them) would serve to “deduct two coins from the dictator’s total earnings.” (The term “punish” was never used.) It was emphasized to participants that their decision only influenced the dictator’s outcome and in no way affected the recipient’s outcome.

Participants were randomly assigned to one of three experimental conditions (14 participants in each condition). In the control condition, participants were instructed that the recipient would not find out about participants’ economic decision. In fact, control participants were told that the recipient was completely unaware that there was a third party in the game. In the privacy breach condition, participants were instructed that the experimenter would be disclosing their economic decision to the recipient, but that the participant’s name would not be disclosed; that is, they could not be identified by the recipient or the dictator. In the anonymity breach condition, participants were instructed that the experimenter would be disclosing both their economic decision and their names to the recipient, but their name would not be disclosed to the dictator; that is, the recipient would be able to identify them, but the dictator could not. It is important to note that for all three conditions, participants were guaranteed anonymity from the dictator and thus did not have to fear extra-laboratory retaliation. Also, all laboratory sessions were carried out identically, aside from the experimental manipulation.

Before participants made their decision, they received a sealed, opaque envelope with the dictator’s decision inside. All participants learned that the dictator kept the entire sum of money (all 10 £1 coins). Past research has shown that people often construe distributions less than half the total amount as violation of a “fairness norm” warranting punishment (e.g., Fehr and Fischbacher, 2004; Pillutla and Murnighan, 1996).

Participants then made their own decisions about how much money they would like to “leave behind” and thus “deduct” from the dictator. To do this, they were instructed to place the amount in an envelope marked “third party” and drop the envelope in a slot box near the door before leaving. Participants were informed that they would not be interacting with the experimenter or anyone else after making their decisions, but instead could leave the building as soon as they were done. This was done to mitigate participants’ concerns about interacting with the experimenter and other players after making their decision. After submitting their envelope, participants completed a brief questionnaire, which asked them to justify their decision and select from a list of emotion terms “any emotions experienced when making their economic decision.” These included the following emotion terms: *anger, anxiety, guilt, shame, excitement, righteousness, fear or worry, admiration, contempt, pride, outrage, envy, or none.*

After completing this form, participants exited the laboratory with whatever amount of coins they decided to keep. This provided participants with a clean get away. Participants did not have to interact with anyone after making their decisions nor did they communicate with other participants between sessions. Participants were later fully debriefed about the study over email.

**Results**

The main dependent measure was the number of coins sacrificed by the participant to punish the selfish dictator. UK£5 was the maximum number of coins that could be sacrificed.

**Figure 1.** Distributions of UK£ used to punish dictator by experimental condition ( $N = 14 \times 3$ )

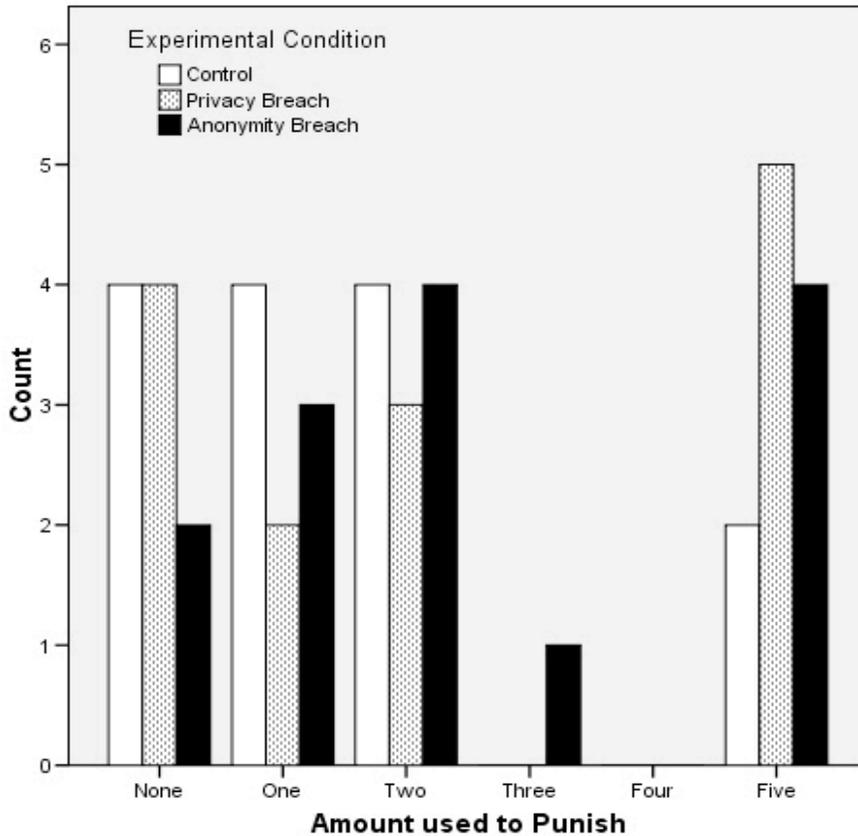


Figure 1 presents the distribution of UK£ used to punish by participants within each condition. Overall, punishment was observed most often in the anonymity breach condition ( $n = 12$ ; 85.7% gave at least £1), although the majority of participants punished the dictator to some extent within the other conditions as well ( $n = 10$ ; 71.4%). Instances of £0 were most frequent in the control (28.6%) and privacy breach conditions (28.6%), while instances of £5 (i.e., full punishment) were observed most often in the privacy (35.7%) and anonymity breach conditions (28.6%).

The mean amount used to punish the dictator in the control condition was 1.57 ( $SD = 1.65$ ); in the privacy breach condition, the mean was 2.36 ( $SD = 2.17$ ); finally, in the anonymity breach condition, the mean was 2.43 ( $SD = 1.87$ ). This pattern of means was in the predicted direction. However, to test our main hypothesis ( $H_A$ : anonymity breach = privacy breach > control), we conducted three planned contrast tests: anonymity vs. privacy, privacy vs. control, and anonymity vs. control. Although comparisons of the experimental conditions against the control condition failed to reach conventional levels of significance ( $t(39) = 1.09, p = .283$ ;  $t(39) = 1.19, p = .242$ ; respectively), differences were

in the direction predicted by our hypothesis ( $H_A$ ) and not the alternative ( $H_B$ : anonymity > privacy = control). As predicted, there was no difference between the anonymity and privacy condition ( $t(39) = .10, p = .922$ ).

Because we predicted that there would be no difference between the experimental conditions, we pooled the data from those conditions and ran a t-test, comparing the combined experimental conditions against the control condition. The results, again, failed to reach conventional levels of significance ( $t(40) = 1.33, p = .191$ ; Levene's test for equality of variance,  $p = .130$ ). The calculated effect size (ES) of the observed difference was .45 (with  $\sigma = 1.82$ ), a "medium" effect size according to Cohen's (1992) ES index, which suggests a possible Type II error.

When punishment behavior was considered as a dichotomous variable (i.e., those who punished [£1-£5] vs. those who did not [£0]), a Chi-square analysis revealed that only in the anonymity breach condition were participants committed to punishing above chance levels ( $\chi^2(1, n = 14) = 7.14, p = .008$ ) ( $ps > .10$  for the privacy breach and control conditions). In other words, participants in the anonymity breach condition consistently punished the dictator, while participants in the other conditions were not consistent.

In terms of gender effects, no specific hypothesis was made. A 3 (condition) x 2 (gender) ANOVA revealed no significant main effect of gender ( $F(1, 36) = .472, p = .496$ ), although a condition x gender interaction approached conventional levels of significance ( $F(2, 36) = 2.75, p = .077$ ). One-way follow-up tests revealed that this nearly significant interaction effect may be explained by a significant difference between male and female participants in the control condition ( $F(1, 12) = 4.91, p = .047$ ), with females ( $M = 2.43, SD = 1.81$ ) punishing to a greater extent than males ( $M = .71, SD = .95$ ).

Participants' written justifications for their economic decisions were coded by both the first author and a second rater naïve to the study's hypotheses (Cohen's  $k = .76$ ; see Table 1 for justification categories and frequencies). Justifications reflected participants' rationales for deducting or not deducting money from the dictator. Some participants provided rationales that related to multiple justification categories. In these cases, multiple codes were assigned. Table 1 is organized by number of coins used to punish the dictator—two or more, one, or none. Punishment behavior was divided into sub-categories of "two or more" and "one" because many participants who gave only £1 to punish felt inclined to justify why they did not give *more* money. Participants who gave "two or more," however, tended to focus on what they did give, rather than what they did not give.

Reasons for punishing included concerns about fairness and justice (90.6%;  $n = 29$ ) and feeling bad for the recipient (9.4%;  $n = 3$ ). (One person invoked a fairness/justice rationale to justify their not punishing, but they were treated as an outlier and therefore omitted.) "Feeling bad for the recipient" justifications only occurred in the anonymity ( $n = 2$ ) and privacy breach ( $n = 1$ ) conditions. Reasons for not punishing (or not punishing more) involved the desire for personal gain (42.9%;  $n = 12$ ), need for money (14.3%;  $n = 4$ ), having nothing to gain from punishing the dictator (14.3%;  $n = 4$ ), or not being able to help the recipient (17.9%;  $n = 5$ ). Finally, justifications from three participants were unscorable or merely restated what the dictator had done without elaboration.

*Anonymity and punishment*

**Table 1.** Justification categories (and frequencies) for punishment behavior ( $N = 42$ )

Justification	Number of coins used to punish dictator		
	Two or more	One	None
Fairness/justice concerns	22	7	0
Feel bad for recipient	3	0	0
Total	25	7	0
Desire for personal gain (or influenced by dictator's desire for personal gain)	4	4	4
Need money	1	1	2
Nothing to gain	0	0	4
Can't help recipient	0	1	4
Unscorable	1	1	1
Total	6	7	15

Table 2 presents participants' reported emotions experienced when making their economic decision. Participants who punished the dictator (i.e., who gave "two or more" or "one") reported experiencing righteousness (28.1%;  $n = 9$ ), anger (21.9%;  $n = 7$ ), contempt (21.9%;  $n = 7$ ), and excitement (21.9%;  $n = 7$ ) more than other emotions. Participants who did not punish (i.e., gave "none") reported experiencing guilt (40.0%;  $n = 4$ ) more than other emotions. However, participants who gave "two or more" or "one" also reported experiencing guilt to a certain extent (18.8%;  $n = 6$ ).

**Table 2.** Emotions reported by participants punishing with two or more, one, or none

	Two or more ( $n = 23$ )	One ( $n = 9$ )	None ( $n = 10$ )	Total ( $N = 42$ )
Anger	5	2	2	9
Anxiety	2	1	2	5
Guilt	3	3	4	10
Shame	2	1	0	3
Excitement	5	2	2	9
Righteousness	5	4	2	11
Fear/worry	2	1	0	3
Admiration	0	1	0	1
Contempt	5	2	2	9
Pride	3	0	0	3
Outrage	3	2	1	6
Envy	1	1	1	3
None	4	0	2	6
Total	40	20	18	78

## **Discussion**

The results from this brief study were generally supportive of our hypothesis that a breach of privacy would promote altruistic punishment in a manner similar to a breach of anonymity due to the reputational consequences it would have been likely to entail in the ancestral past. Participants in the privacy and anonymity breach conditions sacrificed more money to punish an anonymous dictator, who allocated £0 to an anonymous recipient, than participants in the control condition, who were guaranteed privacy and anonymity. Participants in the privacy and anonymity breach conditions sacrificed, on average, half of their endowment to punish, while control participants sacrificed about one-third.

Although contrast tests of these differences failed to reach conventional levels of significance, the pattern of results conformed to our prediction ( $H_A$ : anonymity breach = privacy breach > control), which was inspired by the general evolutionary argument (i.e., the evolutionary legacy hypothesis) that it was difficult—if not impossible—to ensure anonymity in the ancestral past; therefore, the human mind may be biased to respond to a breach of privacy as though it were a breach of anonymity. Although provisional, these findings call into question the conclusions of other investigators who promised their participants anonymity but nevertheless informed these participants that their economic decisions would be made known to other players (e.g., Fehr and Gächter, 2002).

In terms of the consistency at which participants punished, participants in the anonymity breach condition consistently punished the dictator, while participants in the privacy breach and control condition were not consistent. Interestingly, participants in the privacy breach condition appeared to oscillate between sacrificing none (or a trivial amount) and sacrificing all of their money to punish. This suggests perhaps that some participants were responding to the privacy breach cue as though it were a threat to anonymity (or an opportunity to build a reputation), while others were not. We might speculate that some participants were able to consciously override concerns about reputation elicited by the privacy breach cue (by attending to the experimenter's guarantee of anonymity), while others viewed this condition as offering possible opportunities for reputational gain. Future research might investigate this possibility using a cognitive load manipulation.

Despite lower levels of altruistic punishment in the control condition, the majority of control participants did punish to some extent. There are two competing evolutionary explanations for this behavior: First, this behavior may reflect behavioral dispositions for “strong reciprocity” (i.e., dispositions to punish even when there exist no present or future rewards for punishing) that evolved through group selection (see Fehr, Fischbacher, and Gächter, 2002; Gintis, 2000; Gintis et al., 2003). Another possibility is that mechanisms that evolved to maximize individual fitness (e.g., through reputation management) are inadvertently triggered by cues embedded within the experimental context—cues that may be reduced, but never fully eliminated by design parameters (see Burnham and Johnson, 2005). For example, in the context of our third-party punishment game, all participants received information about the existence of two other players (an anonymous dictator and recipient). Despite being guaranteed privacy and anonymity in the control condition, information about the existence of these other players may have been sufficient to trigger concerns about being identified. Future research should certainly investigate this possibility further.

Unexpectedly, male participants in the control condition were significantly less altruistic than female participants. This is consistent with the position that women are more committed to norm enforcement than men (see Eckel and Grossman, 1996b). Another possibility is that men, more than women, require incentives such as social recognition to punish, while women are unilaterally concerned about norm conformity (see Iredale, Van Vugt, Dunbar, 2008, for evidence pertaining to giving behavior). Since there was no difference between men and women in the other conditions, our findings suggest the latter explanation to be more plausible.

Participants who punished the dictator primarily justified their actions in terms of violations of “fairness” and concerns about “justice.” For example, a typical justification for punishing was as follows:

“The unfairness of the [dictator’s] decision affected my choice. Had they distributed the money evenly I would not have taken any of their money away and kept all of my money. However because they wanted to keep all the money and not give the recipient any, I did not mind sacrificing some of my money to detract money from their sum. Were it not for my own self interest I would have left all my money to leave the [dictator] with no money” (control condition)

The above quotation also highlights the dominant justification for not punishing (or not punishing more than one could have), namely, the desire for personal gain (“Were it not for my own self interest...”). Also, a number of participants reported not punishing because such action would not benefit the recipient in any manner. It is impossible to determine whether any justifications in the present study, however, were truly motivational or merely represent post hoc justifications for punishment decisions, since they were not experimentally manipulated.

Interestingly, although very rare, it was only in the anonymity and privacy breach conditions where participants reported “feeling bad for the recipient” as justification for punishing the dictator. One possibility is that by broadcasting one’s sympathy for the “victim” of this social norm violation these participants were capitalizing on the opportunity to communicate their prosocial posture to the recipient. However, because this type of justification was uncommonly reported, we can only speculate. Future studies could examine this possibility further by informing participants that their justifications will be transmitted to the recipient, in addition to their economic decisions.

Participants were also asked to nominate from a list of emotion terms those emotions they had experienced when making their economic decision. Limitations of this emotion-word task include the fact that reported emotional states may be post-hoc reconstructions or causally epiphenomenal, and also that the list was predetermined and may therefore have failed to capture other emotions experienced (e.g., “disgust,” “annoyance”).

Despite these important caveats, the negative social emotions of anger/outrage and positive social emotions of pride/righteousness were, as predicted, reported by participants who punished. These data on affect are important because emotions are generally considered by evolutionary scholars to be the proximate drivers behind adaptive behaviors (see Cosmides and Tooby, 2000; Fessler and Haley, 2003; Frank, 1988). Our findings are in line with appraisal theorists who argue that anger/outrage arises from perceptions of a negative event or transgression as certain, under human control, and brought about by someone external to self (Lerner and Keltner, 2000; Smith and Ellsworth, 1985), and that

pride/righteousness arises from perceptions of being morally responsible and having chosen the socially appropriate course of action (Fessler and Haley, 2003; Tracy and Robins, 2004). In the present study, the dictator's transgressive behavior was presented as certain (i.e., there was no doubt that the dictator was the cause of the event) and deliberate (i.e., the dictator could have chosen another course of action), which satisfies the appraisal criteria for anger/outrage. Participants who punished were responsible for their decision, which was perceived by many to be the "right" course of action—thus, satisfying the appraisal criteria for pride/righteousness.

Emotion categories of excitement and contempt were also reported by participants who punished. This was not predicted, but may be explained by the fact that in Western society "contempt" shares a taxonomic structure with "anger" (i.e., negative/ interpersonal) and "excitement" shares a taxonomic structure with "pride" (i.e., positive/ active/ reward) (see Smith and Ellsworth, 1985; Storm and Storm, 1987).

Participants who did not punish (or who punished very little) reported experiencing guilt more than any other emotion (although whether guilt is an emotion is debated; see Elison, 2005), as predicted; however, against expectations, shame was not reported by non-punishers. One possibility is that the conditions for shame were not present in our study. According to Tracy and Robins' (2004, 2006) appraisal model, guilt and shame both arise from an internal attribution of responsibility for a negative event or a failure to live up to social standards (e.g., committing a transgression or lacking desirable qualities). Guilt and shame, however, differ in that guilt arises from perceptions of the event as unstable, controllable, and incongruent with global aspects of the self, while shame arises from perceptions of the event as stable, uncontrollable, and congruent with global aspects of the self (see also Tangney, 1995; Tangney et al., 2007). If this is correct, we might speculate that non-punishers failed to report shame because they perceived their decisions to be strategic (i.e., controllable) and not emerging from a stable, enduring characteristic of the self (e.g., the self as "selfish" or "undesirable"). Future studies could test this possibility more systematically by comparing the justifications of guilt-prone and shame-prone individuals' for punishing or not punishing.

It is important to point out, of course, that unlike in the ancestral past, contemporary industrial society provides mediated-communication technologies (such as telephones, radios and computers) that make possible a specific type of anonymity—"visual anonymity"—that was not present in the ancestral past. Visual anonymity entails being invisible to one's interactive partner (or having one's interactive partner invisible) during social exchange (see Joinson, 2003). Visual anonymity entails invisibility, but it does not necessarily entail anonymity (in the sense of "not being identifiable"). When people communicate over the telephone or computer (for example, by Instant Relay Chat), they do so under visual anonymity. They cannot be seen, but they may be identified (e.g., by their voice, or deliberately disclose their name). Future research should investigate the extent to which evolved human psychological systems are co-opted by (or respond to) these social artifacts of contemporary technologies, as well as determine the fitness value of such responses in these evolutionarily novel contexts.

**Acknowledgements:** The authors are grateful to Greg Shalkoski for help with recruiting participants and also Herbert Gintis and two anonymous reviewers for helpful comments on

an earlier version of this manuscript.

**Received 28 March 2008; Final revision submitted 8 September 2008; Accepted 8 September 2008**

## References

- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior, 27*, 325-344.
- Barclay, P., and Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings from the Royal Society, B, 272*, 749-753.
- Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 12*, 412-414.
- Baumeister, R.F., Stillwell, A.M., and Heatherton, T.F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin, 115*, 243-267.
- Bohnet, I., and Frey, B.S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior and Organization, 38*, 43-57.
- Boyd, R., Gintis, H., Bowles, S., and Richerson, P.J. (2003). The evolution of altruistic punishment. *Proceedings from the National Academy of Sciences of the United States of America, 100*, 3531-3535.
- Burnham, T.C. (2003). Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior and Organization, 50*, 133-144.
- Burnham, T.C., and Johnson, D.D.P. (2005). The biological and evolutionary logic of human cooperation. *Analyse and Kritik, 27*, 113-135.
- Burnham, T., and Hare, B. (2007). Engineering cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature, 18*, 88-108.
- Chiou, W. (2007). Adolescents' reply intent for sexual disclosure in cyberspace: Gender differences and effects of anonymity and topic intimacy. *Cyber Psychology and Behavior, 10*, 725-728.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Cosmides, L., and Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis and J. M. Haviland-Jones (Eds.), *Handbook of Emotions, 2<sup>nd</sup> ed* (pp. 91-115). New York: Guilford.
- Dawes, C.T., Folwer, J.H., Johnson, T., McElreath, R., and Smirnov, O. (2007). Egalitarian motives in humans. *Nature, 446*, 794-796.
- Eckel, C.C., and Grossman, P.J. (1996a). Altruism in anonymous dictator games. *Games and Economic Behavior, 16*, 181-191.
- Eckel, C.C., and Grossman, P.J. (1996b). The relative price of fairness: Gender differences in a punishment game. *Journal of Economic Behavior and Organization, 30*, 143-158.
- Elison, J. (2005). Shame and guilt: A hundred years of apples and oranges. *New Ideas in Psychology, 23*, 5-32.
- Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63-87.
- Fehr, E., Fischbacher, U., and Gächter, S. (2002). Strong reciprocity, human cooperation

- and the enforcement of social norms. *Human Nature*, 13, 1-25.
- Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.
- Fessler, D.M.T., and Haley, K.J. (2003). The strategy of affect: Emotions in human cooperation. In P. Hammerstein (Ed.), *The genetic and cultural evolution of cooperation* (pp. 7-36). Dahlem Workshop Report. Cambridge, MA: MIT Press.
- Frank, R.H. (1988). *Passions within reason: The strategic role of the emotions*. New York: W. W. Norton and Company.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169-179.
- Gintis, H., Bowles, S., Boyd, R., and Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24, 153-172.
- Haley, K.J., and Fessler, D.M.T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26, 245-256.
- Hardy, C., and Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, 32, 1402-1413.
- Henrich, J., and 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.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science*, 312, 1767-1770.
- Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, 86, 653-660.
- Iredale, W., Van Vugt, M., and Dunbar, R. (2008). Showing off in humans: Male enerosity as a mating signal. *Evolutionary Psychology*, 6, 386-392.
- Joinson, A.N. (2003). *Understanding the psychology of internet behaviour: Virtual worlds, real lives*. Basingstoke: Palgrave Macmillan.
- Kurzban, R. (2001). The social psychophysics of cooperation: Nonverbal communication in a public goods game. *Journal of Nonverbal Behavior*, 25, 241-259.
- Kurzban, R., DeScioli, P., O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28, 75-84.
- Lerner, J.S., and Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgment and choice. *Cognition and Emotion*, 14, 473-493.
- Milinski, M., Semmann, D., and Krambeck, H.-J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature*, 415, 424-426.
- Miller, G. (2000). *The mating mind: How sexual choice shaped the evolution of human nature*. London: Vintage.
- Nowak, M.A., and Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393, 573-577.
- Nowak, M.A., and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291-1298.
- Panchanathan, K., and Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432, 499-502.
- Pillutla, M.M., and Murnighan, J.K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, 68, 208-224.
- Semmann, D., Krambeck, H.J., and Milinski, M. (2004). Strategic investment in reputation.

- Behavioral Ecology and Sociobiology*, 56, 248-252.
- Silke, A. (2003). Deindividuation, anonymity, and violence: Findings from Northern Ireland. *The Journal of Social Psychology*, 143, 493-499.
- Smith, C.A., and Ellsworth, P.C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813-838.
- Smith, E.A., Bliege Bird, R., Bird, D.W. (2003). The benefits of costly signaling: Meriam turtle hunters. *Behavioral Ecology*, 1, 116-126.
- Storm, C., and Storm, T. (1987). A taxonomic study of the vocabulary of emotions. *Journal of Personality and Social Psychology*, 53, 805-816.
- Tangney, J.P. (1995). Recent advances in the empirical study of shame and guilt. *The American Behavioral Scientist*, 38, 1132-1145.
- Tangney, J.P., Stuewig, J., and Mashek, D.J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, 58, 345-372.
- Tracy, J.L., and Robins, R.W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry*, 15, 103-125.
- Tracy, J.L., and Robins, R.W. (2006). Appraisal antecedents of shame and guilt: Support for a theoretical model. *Personality and Social Psychology Bulletin*, 32, 1339-1351.
- Wedekind, C., and Braithwaite, V.A. (2002). The long-term benefits of human generosity in indirect reciprocity. *Current Biology*, 12, 1012-1015.
- Wedekind, C., and Milinski, M. (2000). Cooperation through image scoring in humans. *Science*, 288, 850-852.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110-116.