Emotion expression in human punishment behavior

Erte Xiao* and Daniel Houser*†

Interdisciplinary Center for Economic Science and Department of Economics, George Mason University, 4400 University Boulevard, MSN 1B2, Fairfax, VA 22030

Communicated by Vernon L. Smith, George Mason University, Fairfax, VA, March 24, 2005 (received for review February 14, 2005)

Evolutionary theory reveals that punishment is effective in promoting cooperation and maintaining social norms. Although it is accepted that emotions are connected to punishment decisions, there remains substantial debate over why humans use costly punishment. Here we show experimentally that constraints on emotion expression can increase the use of costly punishment. We report data from ultimatum games, where a proposer offers a division of a sum of money and a responder decides whether to accept the split, or reject and leave both players with nothing. Compared with the treatment in which expressing emotions directly to proposers is prohibited, rejection of unfair offers is significantly less frequent when responders can convey their feelings to the proposer concurrently with their decisions. These data support the view that costly punishment might itself be used to express negative emotions and suggest that future studies will benefit by recognizing that human demand for emotion expression can have significant behavioral consequences in social environments, including families, courts, companies, and markets.

cooperation | ultimatum game | sanction | behavioral economics

Emotion is related to many aspects of social life, from physical survival to social relationships and reproduction (1, 2). With or without self-awareness, humans often display their feelings in different ways when aroused (3–5). However, in many naturally occurring social situations, individuals might believe that it is improper or impossible to reveal their feelings directly to, for example, a perceived antagonist. For instance, a sales clerk might find it improper to confront her customer (6). Because individuals often have a desire to express their emotions, the presence of constraints on expression can have important consequences for human behaviors (7, 8). This research uses ultimatum games to investigate links between constraints on emotion expression (EE) and punishment decisions.

The ultimatum game (9) is widely used to study costly punishment. In this game, one subject (the proposer) starts with, say, $20, and the other subject (the responder) begins with nothing. The proposer suggests a division of the $20 between them, and the responder decides whether to accept the proposed split. If accepted, then the money is split as proposed; if not, then both subjects earn nothing. Consequently, an income-maximizing responder should accept any positive offer, and an income-maximizing proposer would offer the responder the smallest possible positive amount.

In fact, decades of data from ultimatum games show that responders who are offered 20% or so of the total amount choose to reject about half the time (10), and rejection rates increase as responder shares become smaller. Reasons for rejections have been a source of much debate. Recently however, brain imaging data has been collected while responders make their decisions, and the findings suggest that emotions are tightly connected to rejections (11). (For more general research on the link between emotions and costly punishment and debate on the reasons for costly punishment, see refs. 11–17.)

Evolution has likely programmed human responders to prefer to make their negative emotions about unfair offers known to proposers (2, 3). However, standard ultimatum game protocols ensure that responders are constrained from conveying their feelings to proposers in any way except perhaps through choosing costly punishment. It follows that this constraint on expressing emotions could increase the likelihood that responders choose to punish proposers who make unfair offers. Our hypothesis is that responders are less likely to choose costly punishment and correspondingly more likely to accept unfair outcomes when their feelings about unfair offers can be conveyed to proposers in an alternative and less expensive way.

Ultimatum Games with EE

To test our hypothesis we conducted two treatments with the ultimatum game: EE and no EE (NEE). NEE is the standard ultimatum game in which the proposer and the responder are given $20 to split. The proposer decides how many cents of each dollar to keep, and the responder decides whether to accept the offer (divide $20), or to reject the offer (divide $0). In this treatment, rejecting or accepting the offer is the only way for the responder to display a reaction to her proposer about the offer.

The EE treatment is exactly the same as NEE except that the responder is given an opportunity, not a requirement, to write a message to the proposer at no pecuniary cost. Any message is delivered to the proposer concurrently with the responder's decision. Messages cannot have any strategic implications because (*i*) the proposers have made their decisions before they see responders' messages; (*ii*) all experiments take place anonymously, and (*iii*) each pair of subjects plays the game only once. Rather, a message provides an opportunity for a responder to voluntarily display her feelings regarding her proposer's division decision. Our hypothesis is supported if responders in EE use written messages to express emotions and reject unfair offers less frequently than in the NEE treatment.

Experimental Design and Procedures

EE and NEE Treatments. We obtained observations on 296 undergraduates: 62 pairs of subjects in the NEE treatment and 86 pairs in the EE treatment. Experiments included undergraduate students recruited from the general student population at George Mason University by using standard procedures in place at the Interdisciplinary Center for Economic Science. We ran 16 sessions, and the amount to be split in all cases was $20. Subjects were randomly and separately assigned to two rooms, one for proposers and the other for responders. [In the instructions (*Supporting Methods*, which is published as supporting information on the PNAS web site), which closely follow a format used by others (18), we called the proposer "Divider" and responder "Designator."] Each subject was randomly assigned a letter as his or her ID in the experiment. The proposer and responder who received the same letter became a pair. In each room, subjects received an instruction sheet that explained the rules of the game. After reading the instructions, each subject was required to successfully complete a quiz to verify comprehension. The game started after every subject finished the quiz.

First, the proposer indicated his or her proposed split (how many cents of each dollar would go to the proposer and how

Abbreviations: EE, emotion expression; NEE, no EE.

*E.X. and D.H. contributed equally to this work.

†To whom correspondence should be addressed. E-mail: dhouser@gmu.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0502399102

Table 1. Distribution of proposers' offers and responders' messages

	NEE		EE					
					Responders who sent message, %			
Offer	*n*	%	*n*	%	Positive emotion	Negative emotion	Neutral	Total
Responder offered ≥50%					80.56	0	11.11	91.67
20/80	1	1.61	0	0.00				
40/60	1	1.61	4	4.65				
50/50	21	33.87	32	37.21				
Responder offered 40%					22.58	32.26	25.81	80.65
60/40	19	30.65	31	36.05				
Responder offered <40%					0	78.95	10.53	89.48
80/20	14	22.58	15	17.44				
90/10	6	9.68	4	4.65				
Total	62		86					87.21

Proposers' offers are denoted *X*/*Y*, where *X* is proposer's percentage share and *Y* is responder's percentage share. Messages are classified according to the evaluations of 10 objective and hypothesis-blind evaluators.

many would go to the responder) on a decision sheet. After all proposers had finished, the experimenter took all of the decision cards to the responders' room and gave each responder his or her proposer's decision card. The responders decided whether to divide $20 (accept the offer) or $0 (reject the offer). Subjects were given pen and paper in both treatments. In the EE treatment, the responder also received a card for writing a message to her proposer. This card was distributed immediately before the distribution of the proposer's decisions, and messages could have been written before, after, or during their accept/reject decision process. Responders were asked to avoid indecent language but were otherwise given no guidance regarding what, or whether, to write. After responders had finished, the experimenter collected the decision cards (and any message cards in the EE treatment) and returned them to the proposers.

Each pair of subjects played the game once. In both treatments, subjects were given as much time as they liked to make their decisions, and the average length of the two treatments was the same. Subjects were paid privately with cash at the end of the experiment. Each subject received $5 show up bonus in addition to money earned in the game. Subjects were in the lab ≈45 min and earned about $12 total on average.

Message Evaluation. Testing our hypothesis required evaluating the emotional content of our responders' messages. To perform such an evaluation, we used standard Interdisciplinary Center for Economic Science procedures to recruit 10 message evaluators from the general undergraduate population at George Mason University. Potential evaluators were excluded if they had previously participated in any ultimatum game experiment. After being seated in the laboratory, each evaluator was given the responder's instructions from the NEE treatment. We provided evaluators with these instructions because some messages were not necessarily comprehensible without this context. After completing the instructions, the evaluators were given a randomly ordered listing of all 75 anonymous messages written by the responders in the NEE treatment. Subjects were asked to classify the messages as showing positive or negative emotion or as being "neutral" (not positive and not negative). Evaluators were not given any information regarding the situation of the responder who wrote the message: They did not know the proposed split or the responder's decision. Subjects were paid $5 for attending and an additional $5 for completing the entire evaluation. To increase the subjects' attentiveness, they were told that after all evaluations were complete three messages would be randomly chosen as payoff messages. If the subject's evaluation matched the most popular evaluation for a message, then they earned an additional $5. Subjects were in the laboratory for ≈1 hour, and median earnings were $25.

Messages are classified according to the most popular classification chosen by the evaluators. There was a single most popular classification in 71 of 75 cases. The four ties were broken by the investigators' own evaluations.

Results

EE and Punishment. Table 1 describes the distribution of proposers' offers. In both treatments, approximately two-thirds of proposers offer at least 40% of the total amount to the responders and approximately one-third offer 20% or less. EE proposers were aware that responders could send messages along with their accept/reject decisions, but this did not change proposers' decisions in relation to the baseline NEE case: Differences between the two treatments' distributions are not statistically significant (Kolmogorov–Smirnov two-sample test, $P = 0.80$).

In support of our hypothesis, Table 1 also reports that subjects did send messages in EE and that these messages do express emotion (see also Table 2, discussed below). About 87% of all responders wrote a message to their proposer, most of which express emotion. Of the 19 responders who received allocations of 20% or less, 15 (79%) wrote a message expressing a negative emotion, and none expressed a positive emotion. When offered at least half of the total amount, 29 of the 36 responders (81%) displayed positive emotions, and none expressed negative emotions, which is not surprising given that the responders in both treatments accepted all offers that allocated at least $10 of the $20 to them.

Rejection rates differ between the two treatments when the proposer offers the responder $4 (20%) or less. In the baseline NEE case, 60% (12 of 20) of such offers are rejected, a frequency that lines up well with previously reported results (10). However, in the EE treatment, only 32% (6 of 19) reject the unfair offer, and this difference is statistically significant (Mann–Whitney test, $z = 1.757$; one-tailed, $P = 0.04$). Inspection of Table 1 reveals that most of the data are in cases in which the responder is offered 20% ($4). This 20% offer occurs 14 times in NEE, with seven responders (50%) choosing to reject. In contrast, only 3 of 15 responders (20%) do so in EE, and this difference is statistically significant (Mann–Whitney test, $z = 1.669$; one-tailed, $P = 0.05$). (Because the 90%/10% choice is made very infrequently by proposers, it is not possible to draw inferences based on responder decisions in that cell alone.) Finally, note

Table 2. Messages written when the offer is 80/20 or 90/10

Offer	Subject	Accept	Message	Emotion
80/20	1	No	Sorry, I'm a person too. When the cards are all in my hand, you should try to appease me instead of offend me. There was a 50/50 split. It couldn't have been easier. So, since you decided you are obviously better than I am. You get nothing. Enjoy it, I know I will.	Negative
	2	No	If you would have been less greedy than maybe we would have gotten some money. Treat everyone as you want them to treat you.	Negative
	3	No	Should not have been greedy. Oh well, you make nothing.	Negative
	4	Yes		
	5	Yes	Thanks For Nothing.	Negative
	6	Yes	It would have been better if you had chosen D. I was going to split $0, so you would no gain. But $4 is better than nothing. So I decide to go with it. Have I been the divider, I would have chosen D.	Negative
	7	Yes	I guess I'll do $20. You are getting way more than me. But if I screw you over I get no Money either :-)	Negative
	8	Yes	Tuesday is election day! Vote for Kerry Read the platform johnkerry.com :-)	Neutral
	9	Yes	Not fair, I wish I am the divider! But I get $4 is better than none.	Negative
	10	Yes	Too selfish. I would rather get nothing and let you get a penny.	Negative
	11	Yes	Dude, that's kina greedy and I'm seriously contemplating designating $0. . . I was hoping you'd choose D so we'd both be happy but whatever, Grrrr. . .	Negative
	12	Yes	You suck, you are lucky I'm broke! If you did the A I would have put 0.	Negative
	13	Yes	You are lucky I'm broke!	Negative
	14	Yes		
	15	Yes	I should have chosen to divide by $0 but I'll take the $9 since I don't like wasting my time. Enjoy your $16.	Negative
90/10	1	No	Hey, we could both benefited equally, but no -sorry.	Negative
	2	No	I don't think so buddy!	Negative
	3	No	Well, we all want to make a little bit. I have the money here. Since you are my divider, I think it would better for the both of us to go for rule D. either that or we won't get nothing at all. . .	Neutral
	4	Yes	We should have divided the money equally. Don't be so greedy. People are always out for themselves	Negative

In the Offer column, the first number is the proposer's percentage share and the second number is the responder's percentage share. As described in the experiment's instructions (*Supporting Methods*), ''D'' stands for an equal-split offer and ''A'' stands for an offer of 10% to the responder (so 90% to the proposer). The last column shows whether the message is classified as expressing positive or negative emotion or is neutral (i.e. neither positive nor negative).

that rejection rates in the 60%/40% cell are ≈10% in each treatment. Fig. 1 summarizes these results.

Responders' Messages. Table 2 details the messages written by all responders who faced offers of 20% or less, whether the message was classified as expressing a positive or negative emotion or was neutral (expressing neither a positive nor negative emotion), and each responder's decision (all the messages are provided in Table 3, which is published as supporting information on the PNAS web site). When the offer is exactly 20%, 10 of 12 responders who accepted the offer wrote a message, and 9 of these 10 messages were classified by the evaluators as expressing negative emotions. In addition, note that some responders accepted unfair offers even while indicating that they should not, which might indicate that the egalitarian or retaliation motivation (16, 17) for punishment can be diminished by providing subjects an opportunity to express their feelings. Overall, these data provide convergent support for the possibility that costly punishment is used by responders to express emotions and that responders are less likely to use costly punishment and instead accept unfair outcomes if they have a less expensive alternative mechanism to express negative emotions toward the proposers.

Discussion

Negative emotions like anger or disapproval can be triggered when individuals are treated unfairly (11–15), and evolutionary theory argues there can be benefits to expressing negative emotions in some contexts (2, 3, 19, 20). (More general discussion of the evolutionary importance of punishment is provided in refs. 21–23.) In our EE treatment, for example, responders might feel better after explicitly displaying their emotions to proposers (4, 24); or, perhaps sending messages of disapproval directly to one's proposer might be a satisfying alternative form of punishment (13, 25); or, a responder might believe that accepting a low offer would be interpreted by the proposer as

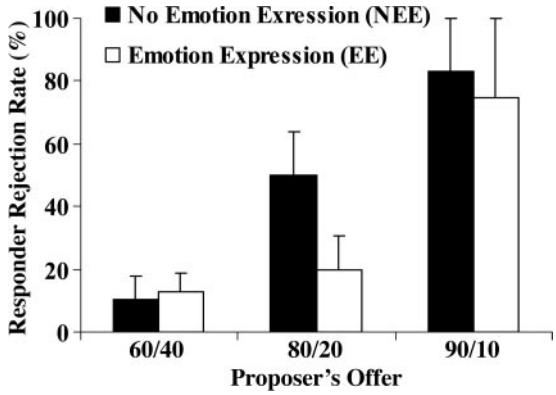

Fig. 1. Rejection rates when responders are offered <50%. Columns show mean ± SEM. Proposers' offers are denoted *X*/*Y*, where *X* is proposer's percentage share and *Y* is responder's percentage share. When proposers offer 60/40, EE responders reject the offer at nearly the same rate as NEE. When the offer is either 80/20 or 90/10, the EE rejection rate is lower. The difference is statistically significant in the 80/20 cell, as well as when the 80/20 and 90/10 data are pooled ($P < 0.05$). A responder is more likely to accept an unfair division if she can express her emotions about the offer concurrently with her decision.

indicating that the responder accepts an inferior position. By expressing anger or disapproval regarding the low offer, responders can deny this interpretation.

When direct channels for expressing emotions are either impossible or undesirable, our results suggest that humans might instead resort to indirect or even costly methods to convey negative feelings, particularly costly punishment. Our results highlight the importance humans attach to expressing negative emotions.

Constraints on expressing emotions might be a contributing factor for decisions typically observed in many naturally occurring and experimental environments, including highly studied trust, public goods, and bargaining games (18, 26–33). For example, subjects in public goods games are generally found to decrease their contribution to the public goods when others contribute little (28–32). If these decisions are partially motivated by a desire to express unhappiness to free riders, then such reductions in contributions might be less common if subjects were provided an alternative way to express their feelings.

In addition to expressing negative emotions, it is important to emphasize that ≈80% of responders in our experiments displayed positive emotions toward proposers when they received fair offers. Presumably, a demand to express positive emotions can also affect decisions. For example, in a typical trust game (26, 27), where the investor transfers part of her endowment to a trustee, the only way for the trustee to express gratitude is to reciprocate and return some amount to the investor. If this reciprocity is in fact motivated by human demand to express positive emotions (such as happiness or appreciation), then measured trustworthiness (amounts returned to investors by trustees) might decrease if trustees are given an alternative, less-costly channel to express appreciation to investors. Further exploration in this area, particularly efforts at eliciting the "demand curve" for expressing positive and negative emotions, would be useful.

Our results rely in part on classification of the emotional content of responders' messages. The classification approach we adopted is standard in its use of independent, objective, and hypothesis-blind human evaluators. Nevertheless, we cannot know the "true" emotion behind any of the messages we collected. Having said this, it should be reiterated that there was substantial agreement among our independent evaluators with respect to the emotional content of the vast majority (95%) of responders' messages.

The desire to express emotions, and constraints on that demand, are a ubiquitous feature human social interaction. The results of our study are a step toward an improved understanding of human behavior in environments that involve emotions (6, 34, 35). Our research, of course, provides only one perspective on how emotion is connected to human behavior. Emotions might have different effects in different contexts. More work and specific models are needed to advance our understanding of how emotions are involved in human decision-making processes.

We thank two referees for useful comments and Tyler Cowen, Ernst Fehr, Timothy Ketelaar, Robert Kurzban, Francisco Parisi, Vernon Smith, and Bart Wilson for valuable thoughts on this project. This work was supported by the International Foundation for Research in Experimental Economics and the National Science Foundation.

1. Keltner, D. (2003) *Ann. N.Y. Acad. Sci.* **1000,** 222–243.
2. Cosmides, L. & Tooby, J. (2000) in *Handbook of Emotions*, eds. Lewis, M. & Haviland-Jones, J. M. (Guilford, New York), pp. 91–115.
3. Darwin, C. & Ekman, P. (1998) *The Expression of the Emotions in Man and Animals* (Oxford Univ. Press, Oxford).
4. Kennedy-Moore, E. & Watson, J. C. (1999) *Expressing Emotion: Myths, Realities, and Therapeutic Strategies* (Guilford, New York).
5. Marshall, J. R. (1972) *Arch. Gen. Psychiatr.* **27,** 786–790.
6. Rafaeli, A. & Sutton, R. I. (1989) in *Research in Organizational Behavior*, eds. Cummings, L. L. & Staw, B. M. (JAI, Greenwich, CT), Vol. 11, pp. 1–43.
7. Gross, J. J. & Levenson, R. W. (1993) *J. Pers. Soc. Psychol.* **64,** 970–986.
8. Gross, J. J. (1999) *Cognit. Emotion* **13,** 551–573.
9. Güth, W., Schmittberger, R. & Schwarze, B. (1982) *J. Econ. Behav. Organ.* **3,** 367–388.
10. Camerer, C. (2003) *Behavioral Game Theory: Experiments in Strategic Interaction.* (Princeton Univ. Press, Princeton, NJ).
11. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. (2003) *Science* **300,** 1755–1758.
12. Fehr, E. & Gächter, S. (2002) *Nature* **415,** 137–140.
13. Elster, J. (1999) *Alchemies of the Mind: Rationality and the Emotions* (Cambridge Univ. Press, Cambridge, U.K.).
14. Pillutla, M. M. & Murnighan, J. K. (1996) *Organ. Behav. Hum. Dec.* **68,** 208–224.
15. Camerer, C. (2003) *Science* **300,** 1673–1675.
16. Fowler, J. H., Johnson, T. & Smirnov, O. (2005) *Nature* **433,** E1.
17. Fehr, E. & Gächter, S. (2005) *Nature* **433,** E1–E2.
18. Andreoni, J., Castillo, M. & Petrie, R. (2003) *Am. Econ. Rev.* **93,** 672–685.
19. Frank, R. H. (2002) in *Evolution and the Capacity for Commitment*, ed. Nesse, R. M. (Russell Sage Found., New York), pp. 57–76.
20. Hirshleifer, J. (1987) in *The Latest on the Best: Essays on Evolution and Optimality*, ed. Dupré, J. (Bradford Books–MIT Press, Boston), pp. 307–326.
21. Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 3531–3535.
22. Boyd, R. & Richerson, P. J. (1992) *Ethol. Sociobiol.* **13,** 171–195.
23. Axelrod, R. & Hailton, W. D. (1981) *Science* **211,** 1390–1396.
24. Pennebaker, J. W. (1997) *Opening Up: The Healing Power of Expressing Emotion* (Guilford, New York).
25. Masclet, D., Noussair, C., Tucker, S. & Villeval, M. C. (2003) *Am. Econ. Rev.* **93,** 366–380.
26. Camerer, C. & Weigelt, K. (1988) *Econometrica* **56,** 1–36.
27. Berg, J., Dickhaut, J. & McCabe, K. (1995) *Games Econ. Behav.* **10,** 122–142.
28. Fehr, E. & Gächter, S. (2000) *Am. Econ. Rev.* **90**, 980–994.
29. Kurzban, R. & Houser, D. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 1803–1807.
30. Houser, D. & Kurzban, R. (2002) *Am. Econ. Rev.* **92,** 1062–1069.
31. Andreoni, J. (1990) *Econ. J.* **100,** 464–477.
32. Isaac, M. & Walker, J. (1988) *Q. J. Econ.* **103,** 179–199.
33. McCabe, K. A., Rassenti, S. J. & Smith, V. L. (1998) *Games Econ. Behav.* **24,** 10–24.
34. Lind, E. A. & Tyler, T. R. (1988) *The Social Psychology of Procedural Justice (Critical Issues in Social Justice)* (Plenum, New York).
35. Umbreit, M. S. (1995) *Mediating Interpersonal Conflicts: A Pathway to Peace* (Erickson Mediation Inst., Minneapolis).