



Inequality-Seeking Punishment

Daniel Houser and Erte Xiao

2009

Discussion Paper

Interdisciplinary Center for Economic Science
4400 University Drive, MSN 1B2, Fairfax, VA 22030
Tel: +1-703-993-4850 Fax: +1-703-993-4851
ICES Website: www.ices-gmu.org
ICES RePEc Archive Online at: <http://edirc.repec.org/data/icgmuus.html>

Inequality-Seeking Punishment

Daniel Houser

Department of Economics & Interdisciplinary Center for Economic Science
George Mason University

Erte Xiao

Department of Social and Decision Science
Carnegie Mellon University

Abstract: Inequality aversion is a key motive for punishment, with many prominent studies suggesting people use punishment to reduce or eliminate inequality. Punishment in laboratory games, however, is nearly always designed to promote equality (e.g., rejections in standard ultimatum games) and the marginal cost of punishment is typically non-trivially positive. As a consequence, individual preferences over punishment outcomes remain largely uninformed. We here report data from a laboratory experiment using dictator games. We find that when people are treated unfairly they systematically prefer to use punishment to create advantageous inequality. Our results shed new light on human preferences over punishment outcomes, and have important implications for the design of mechanisms to deter misconduct.

Acknowledgments: We wish to thank John Duffy, Vernon Smith, Roberto Weber and participants in seminars at the International ESA 2009, North American ESA 2008 and University of Pittsburgh for many useful comments. We are grateful for grants to the first author from the International Foundation for Research in Experimental Economics (IFREE-103) and to the second author from the Russell Sage Foundation.

Introduction

Costly punishment is ubiquitous in both naturally-occurring and designed environments (Fehr and Gächter, 2000). Substantial research has investigated motives for costly punishment, with fairness and inequality aversion receiving attention as important motives for punishment decisions (Falk et al, 2005; Dawes et al. 2007; Cox et al, 2008; Xiao and Houser, 2005)¹. Indeed, inequality aversion is the foundation of a literature of models incorporating other-regarding preferences (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Surprisingly, however, little evidence exists on preferences over punishment outcomes. Of particular importance is whether, following unfair treatment, punishment is used instrumentally to achieve inequality that favors the punisher. This issue is key because inequality-seeking punishment is might likely lead to retaliatory spirals that substantially diminish economic efficiency (Dreber et al., 2008). This paper reports data from an experiment that sheds new light on preferences over punishment outcomes, thus providing new insight on the role of punishment in promoting and maintaining social norms (Fehr and Gächter, 2002).

In most previous studies of punishment, a participant could choose to incur a cost Y in order to deduct an amount $Z > Y$ from a specific other (a ratio of 1:3 is common, e.g., Fehr and Gächter, 2000). Using punishment in this manner can evidently decrease inequality. While decisions in these environments reveal constrained preferences over costly punishment outcomes, they do not readily allow one to infer preferences over outcomes not constrained by costs.

Some evidence hints that people might prefer punishment to result in unequal outcomes. For example, a recent study (Herrmann et al., 2008) found non-trivial amounts

¹ Falk et al (2005) shows punishment can also be motivated by desires for retaliation and spitefulness.

of antisocial punishment in a public goods game. Even so, the complexity of the repeated public goods game makes drawing inferences about punishers' goals difficult². Moreover, a recent survey study by Mocan (2008) draws attention to the prevalence of vengeance. Punishment, particularly when motivated by the perception of negative "intentions," might serve to exact revenge rather than to maintain a norm of equality.

In this paper, we investigate preferences over punishment outcomes using data obtained from a novel two-stage dictator game with punishment. The game setup allows receivers to punish dictators by destroying the dictators' earnings (receivers do not benefit from the destruction). Punishment has a positive fixed cost but zero marginal cost; that is, a receiver can decide whether to pay a small fixed amount for the right to punish at any level.

We consider games where allocations are decided by a person as well as games where allocations are decided by a random device. Consistent with previous results (Blount, 1995), we find that punishment occurs much more frequently when unfair allocations are determined intentionally by a person than when decided randomly. Our new finding is that in both our random and intention treatments, a large majority of the receivers who choose to punish also choose to destroy sufficient dictator earnings to ensure that receivers are left in a position of advantageous inequality. Thus, our evidence supports inequality-seeking punishment.

² For example, in a public goods game, subjects can punish anyone they want. Therefore, any subject can receive punishment from multiple persons. Likewise, punishers know this, which means that the outcome of the income distribution is not decided by any single punisher.

Experiment Design

We study decisions in both intentional and random allocation dictator games. We use both games to explore whether preferences for punishment outcomes vary with unfair intentions. In both treatments, we allowed receivers to punish their respective dictators at a small fixed cost. Subjects were recruited at the University of Pennsylvania using the “Experiments @ Penn” web-based recruitment system. Participants were randomly matched and assigned the role of either dictator or receiver.

I. Intention treatment

In the first stage, the dictator received an initial endowment of \$8. He/she could transfer any even amount (\$0, \$2, \$4, \$6 or \$8) to the receiver. Next, at the beginning of the second stage, both the dictator and the receiver received an additional \$2.

In the second stage, the receiver decided whether to use \$1 to destroy any integral amount of the dictator’s money up to the amount he/she kept in the first stage plus the additional endowment of \$2 (that is, dictator earnings could not be less than zero). We used the strategy methods to elicit receivers’ preferences over punishment outcomes. Thus, the receiver was asked to make this decision for each possible transfer decision the dictator could have made in the first stage. Both the dictator and the receiver were aware that the final payoff would be determined by the dictator’s transfer decision in the first stage and the receiver’s corresponding punishment decision in the second stage.

Note that when a dictator sent less than half of the \$8 endowment to a receiver, \$0 or \$2, she generated self-advantageous inequality. In the game’s second stage, however, receivers had an opportunity to pay \$1 and eliminate or reverse that inequality advantage.

II. Random treatment

The Random treatment was identical to the Intention treatment except that dictators were passive. In the first stage, a computer randomly split \$8 between the dictator and the receiver. The receivers knew this was the case, but continued to have the opportunity to make second stage decisions as described above. Thus, differences in receivers' decisions between treatments inform the extent to which dictators' earnings destruction is due only to inequality aversion as compared to a combination of that and unfair dictator intentions.

Results

We obtained observations on 222 subjects: 48 pairs in the intention treatment and 63 pairs in the random treatment³. Only 5 of 111 receivers deducted a positive amount when dictators made an equal or better-than-equal split for the receiver. Two of these deductions occurred in the intention treatment, while the others occurred in the random treatment. Since our focus is inequality-motivated punishment, we focus on receivers' punishment decisions after being sent \$2 or \$0 out of \$10.

First, about 54% of receivers in the Intention treatment deducted a positive amount from the dictator when the dictator sent either \$2 or \$0. In the Random treatment, significantly fewer did so (only 19%, Z-test, $p < 0.01$). Table 1 reports the average amount receivers deducted from dictators for each possible outcome in each treatment. The amount deducted in the second stage was larger when the amount sent in the first stage was smaller. Also, receivers receiving \$2 or \$0 deducted significantly more on average in the Intention treatment than in the Random treatment (Mann-Whitney, $p < 0.001$).

³ Two dictators in the intention treatment were confused with the experiments. We thus exclude the two in the analysis related to dictators. As we use strategic methods, we still keep all the data for the receivers.

Next, we say that punishment is “inequality-seeking” whenever a receiver is willing to pay a positive cost (\$1) to change an unequal allocation from one that favors the dictator to one that favors the receiver. Thus, a receiver is an “inequality-seeking punisher” when : 1) the receiver deducts \$10 if receiving \$0 in the first stage; or 2) the receiver deducts either \$8, \$7 or \$6 from the dictator if receiving \$2.

In the intention treatment, we find that unequal offers lead to inequality-seeking punishment with frequency 39.6%. In the random treatment, about 12.7% of receivers with unequal initial allocations displayed inequality-seeking punishment behavior, and the difference is statistically significant (Z test, $p < 0.01$). Figure 1 plots the distribution of the amount of dictators’ earnings destroyed by receivers in each treatment. The distributions are statistically significantly different (Kolmogorov-Smirnov test, $p < 0.05$). Both results suggest that inequality-seeking punishment is motivated not only by the unequal distribution of outcomes in the first stage but also by unfair dictator intentions.

In both treatments, two-thirds of punishments are inequality-seeking. In the Intention treatment, 73% of punishers are inequality-seeking (19 out of 26) while in the Random treatment the frequency is $2/3$ (8 out of 12), and these two proportions do not differ statistically (Z test, $p = 0.69$).

In summary, we offer rigorous evidence for inequality-seeking punishment. A large portion of receivers who receive unequal allocations are willing to incur a cost to punish the dictator and create a final allocation that is unequal in favor of themselves.

Discussion

Previous research has shown that people are inequality-averse and that this inequality aversion often motivates costly punishment behavior. However, knowing the motivations of punishment is not enough to allow us predict the outcome of punishment. In this paper, we provide rigorous evidence that disadvantageous inequality motivated punishments are often used by the punisher to achieve a position of advantageous inequality. Consistent with previous research (Blount, 1995; Nelson, 2002), our data suggest people are more willing to tolerate inequality when it is caused by nature than when it is intentionally created by humans. Nevertheless, in both cases, a large majority of punishers attempt to achieve advantageous inequality.

Our finding that people's willingness to pay to obtain advantageous inequality is mediated by intentions is consistent with the finding that people take pleasure in the act of punishing unfair people, or seeing unfair people punished (De Quervain et al., 2004; Singer et al., 2006.) In turn, this is consistent with the possibility that revenge is a key motive for punishment decisions. Since retaliatory punishment can lead to a downward spiral in cooperation and economic efficiency (Dreber et al., 2008), our results emphasize the importance of designing institutions to promote fairness (Xiao and Houser, 2009), while also encouraging tolerance in those cases when inequality unavoidably emerges.

References

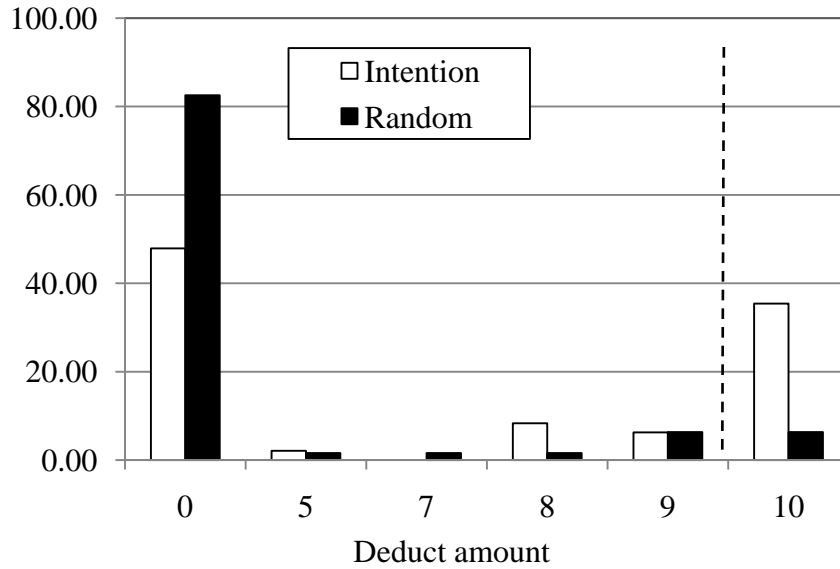
- Blount, S. (1995). When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences. *Organizational Behavior and Human Decision Process* 63 (2), 131-144.
- Bolton, G. and Ockenfels, A. (2000) "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90, 166-193.
- Cox, J., Sadiraj, K. and Sadiraj, V. (2008) "Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior" *Experimental Economics*, 11, 1-24.
- Dawes CT, Fowler JH, Johnson T, McElreath R, Smirnov O. (2007) "Egalitarian Motives in Humans," *Nature* 446 (12), 794-796
- Dreber, A., Rand, D.G., Fudenberg, D. and Nowak, M.A. (2008). "Winners don't punish." *Nature*, vol. 452, 348-351.
- De Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. (2004) The neural basis of altruistic punishment. *Science*, 305, 1254-58.
- Falk, A., Fehr, E. and Fischbacher, U. (2005). "Driving Forces behind Informal Sanctions", *Econometrica*. 2017-2030.
- Fehr, E. and Schmidt, K. (1999) "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 817-868.
- Fehr, E. and Gächter, S. "Altruistic Punishment in Humans " *Nature* 415, 10 January 2002, 137-140
- Fehr, E. and Gächter, S. 2000 ." Cooperation and Punishment in Public Goods Experiments. *American Economic Review*. 90 (4), 980-994.
- Herrmann, B. , Thoni, C., Gächter, S. (2008) "Antisocial Punishment Across Societies", *Science*, 319(5868), 1362-1367.
- Mocan, N. (2008). "Vengeance", NBER working paper No. 14131.
- Nelson W.R. Jr., (2002). Equity and intention: it's the thought that counts. *Journal of Economic Behavior and Organizations*, 48(4), 423-430.
- Singer, B Seymour, JP O'Doherty, KE Stephan, RJ. (2006). "Empathetic neural responses are modulated by the perceived fairness of others." *Nature*, vol. 439, 466-469.
- Xiao, E. and Houser, D. (2005). "Emotion expression in human punishment behavior." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, 7398-7401.
- Xiao, E. and Houser, D. (2009). "Avoiding the sharp tongue: Anticipated written messages promote fair economic exchange." *Journal of Economic Psychology*. 33(3), 393-404.

Table 1. Receiver's punishment decision by treatment

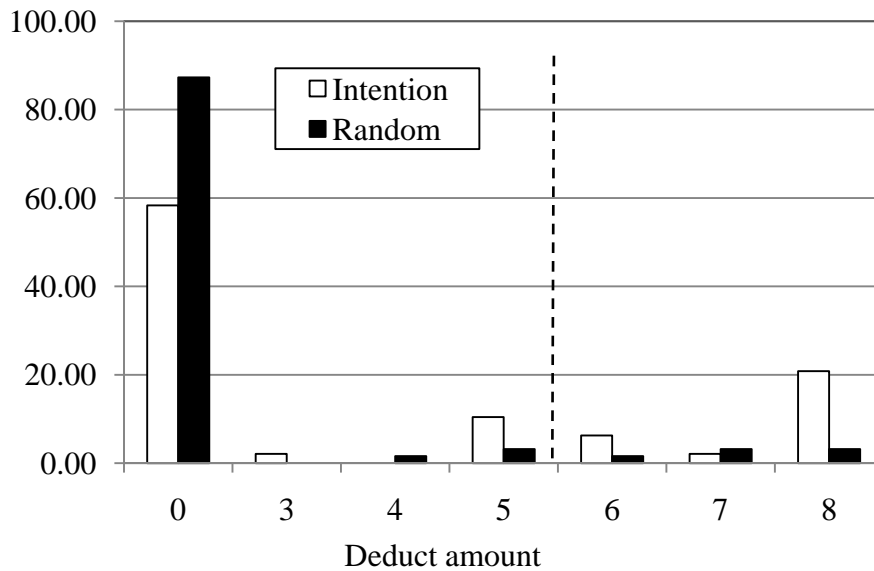
Amount received	Intention (obs.=48)		Random (obs.=63)	
	Mean	S.e.	Mean	s.e.
0	4.88	0.69	1.52	0.43
2	2.77	0.50	0.79	0.27
4	0.25	0.17	0.16	0.11
6	0.08	0.08	0.00	0.00
8	0.00	0.00	0.03	0.03

Figure 1. Distribution of Receivers' Punishment Amounts by Treatment

A. When receivers are allocated \$0 in the first stage.



B. When receivers are allocated \$2 in the first stage.



Note: Left to the dotted line is inequality-seeking deduction.