

The devil you know: The effects of identifiability on punitiveness

Deborah A. Small and George Loewenstein

Carnegie Mellon University

Correspondence Address:

Deborah A. Small
Department of Social & Decision Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
Phone: 412-268-1207
FAX: 412-268-6938
Email: dsmall@andrew.cmu.edu

*We thank Sam Issacharoff (for, among other things, suggesting the title), Jennifer Lerner, Margaret Clark, and Linda Babcock for their help. Small's participation in this research was supported by the Russell Sage Foundation. Loewenstein's participation in this research was supported by the Center for Integrated Study of the Human Dimensions of Global Change, a joint creation of the National Science Foundation (SBR-9521914) and Carnegie Mellon University.

Abstract

Prior research has found that people are more generous toward an identifiable (determinate) victim than toward a statistical (indeterminate) victim. In the present study, we demonstrate an equivalent effect for punitiveness. We find that people are more punitive toward determinate wrongdoers than toward equivalent, but indeterminate, wrongdoers, even when determining the wrongdoer conveys no meaningful information about him or her. To account for the effect of identifiability on both generosity and punitiveness, we propose that affective reactions of any type are stronger toward a determinate than toward an unidentified target. Consistent with such an account, the effect of determinateness on punishing behavior was mediated by anger and blame.

...it is in particular instances only that the propriety or impropriety, the merit or demerit, of actions is very obvious or discernible.When we consider virtue and vice in an abstract and general manner, the qualities by which they excite these several sentiments seem in a great measure to disappear, and the sentiments themselves becomes less obvious and discernable.

Adam Smith, *The Theory of Moral Sentiments*

Past research has shown that human empathy differs reliably toward actual, 'identified', victims on the one hand, and more abstract or 'statistical' victims on the other (Fetherstonhaugh, Slovic, Johnson, & Friedrich, 1997; Small & Loewenstein, 2003). As Schelling (1968) wrote in what may have been the first explicit treatment of the phenomenon, "the death of a particular person invokes "anxiety and sentiment, guilt and awe, responsibility and religion, [but]...most of this awesomeness disappears when we deal with statistical death." Schelling's passage not only identifies the phenomenon, but also proposes a plausible psychological mechanism involving emotions. It suggests that identifiable victims evoke sympathy and a sense of moral responsibility that is lacking in considerations of statistical victims.

In this paper we examine whether the discrepancy in treatment of statistical and identifiable victims noted by Schelling and supported by subsequent research might be a special case of a more general phenomenon that could be termed an *identifiable other* effect whereby *any* identifiable target evokes a stronger emotional and moral reaction than an equivalent, but unidentifiable target. If identifiable targets of any type produce stronger emotional reactions, then identification should also tend to intensify negative

feelings, if these are the dominant emotional reactions to a target. This is the prediction we test in the current paper.

Specifically, we test for an effect of identifiability on punitiveness, adapting a research design borrowed from our earlier work on the identifiable victim effect (Small & Loewenstein, 2003). The original design was intended to get around the problem that identifying a victim generally means providing information about him or her, so it is always possible that any observed increment in empathy toward identifiable victims could be due to the specific information provided about the victim rather than to identifiability *per se*. Our study avoided this problem by identifying victims without providing any information about them, a manipulation that we called “determinateness.”

In one of the earlier studies, we assigned each member of a group of research participants with a number and endowed each with \$10. Based on a drawing of numbers, half – the 'victims' -- were made to return the money. We then gave each of the participants who had retained the \$10 the opportunity to share their money with one of those who had lost their endowment. In the determinate (identifiable) condition, the potential giver *first* drew the number of one victim from a bag, then decided how much to give to that victim (knowing, however, that he/she would never learn the actual identity of the victim). In the indeterminate (unidentifiable) condition, in contrast, the potential giver decided how much to give just *before* drawing the victim's number. Donations were about twice as large, on average, in the determinate condition as in the indeterminate condition, despite the fact that determining the victim provided no information about them. Follow-up research, in which we raised money for the

charitable organization "Habitat for Humanity," revealed a similar effect in a more naturalistic setting.

Beyond generalizing the earlier work beyond reactions to victims, the current study also examines whether any observed differences in the punitiveness exhibited toward determinate and indeterminate perpetrators would be mediated by different affective reactions. Adam Smith's assertion that "when we consider virtue and vice in an abstract and general manner, the qualities by which they excite these several sentiments seem in a great measure to disappear," as well as Schelling's contention that identified victims evoke "anxiety and sentiment, guilt and awe," both reveal an implicit theory that identification matters because it leads to more intense emotional reactions. The work of Sherman and colleagues (Hamilton and Sherman, 1996; Sherman, Beike and Ryalls, 1999) further supports the idea that people use distinct processes to make judgments of specific as opposed to general targets, and specifically the idea that generalities tend to evoke semantic representations, whereas specific instances evoke 'episodic' representations. We contend that the representations of specific, identifiable targets are highly affect-laden, engaging the perceiver at a particularly intense level. But, such a mediating role of affect has not been tested, including in our own prior work, which examined generosity toward statistical and identifiable victims, but did not incorporate measures of affect.

The relevant emotions to examine in the context of punitiveness, we assumed, would be anger and blame. Prior research has show that perceived intentional harm evokes anger (Bentacourt & Blair, 1992) as well as blame (e.g., Shaver, 1985), and that these two emotions interact with one-another. Anger, like sympathy, is a moral emotion,

which can produce strong inferences of blame (Averill, 1983; Weiner, 1995); Blame, likewise, intensifies anger, such that anger and blame have a significant recursive relationship (Quigley, & Tedeschi, 1996). Moreover, reactions of anger and blame naturally induce a desire to punish (Lerner, Goldberg, & Tetlock, 1998; Solomon, 1990). Hence, to the extent that identifiable wrongdoers evoke stronger emotional and moral reactions than unidentifiable wrongdoers, we should expect them to be punished more severely.

Present Study

Experiment Overview

To test the effect of identifiability on punitiveness, we created a situation in which participants who had behaved cooperatively in a social dilemma at their own expense were given the opportunity to penalize another participant who had behaved in a self-interested fashion at the expense of others.¹ Analogous to our earlier study, identifiability was manipulated by having contributors make the decision either just before or just after they had drawn the identification number of a non-contributor. Participants made real decisions about cooperating and punishing which affected their actual payoffs. This is worth noting because of the emotional mechanism that we propose. If choices were merely hypothetical, we would expect less of an effect, or no effect at all, since people generally mispredict their emotions in a hypothetical context (e.g., Van Boven, Loewenstein, Welch, & Dunning, 2003).

We predicted that people would be more likely to punish, at their own expense, an *identified* (determinate) non-contributor more severely than an *unidentified* (indeterminate) non-contributor. Second, when given a choice to penalize a non-

contributor, participants would react with greater anger and blame toward an *identified* non-contributor than toward an *unidentified* non-contributor. Finally, we predicted that the effect of identifiability (determinateness) on punishment would be mediated by feelings of anger and blame.

Method

Participants

One hundred and forty undergraduate and masters' students (58 females and 81 males) at Carnegie Mellon University participated in the study. They received no participation fee other than whatever sum of money they earned from the game. There were no significant gender differences on any measure, so male and female data were combined in all analyses.

Procedure

Participants were recruited in groups of ten. They were seated facing away from one another and were instructed not to speak or turn around and look at one another during the course of the experiment. The experimenter informed the participants that all decisions they made would be anonymous and that, at no point during or after the experiment, would anyone learn the identity of anyone in their group. Participants were told that they would receive their payments from the outcome of the game in sealed envelopes, so that they would learn only about their own payoff from the game.

At the beginning of each experimental session, the experimenter had each participant draw a number from a bag containing pieces of paper labeled with numbers from 1-10; each participant drew a single number. Participants were told that the

experiment would consist of two rounds. Each participant then received the following written instructions for Round 1:

At the beginning of Round 1 you, and every other participant, receive \$5. You and each of the other 9 group members must decide whether to contribute your \$5 to the group or to keep it for yourself. If you contribute the money, then everyone in the group will receive \$1.25 from you. If you do not, then everyone in the group will receive nothing from you. Therefore, your income from the experiment depends on what you do and what everyone else does.

If everyone contributes all of their money, including you, then you will all make \$11.25 (9 x \$1.25).

If everyone keeps their \$5.00 and no one contributes theirs, then everyone will make \$5.00.

The most you can make would be if you keep your money and everyone else contributes, in which case you would make \$16.25.

The least you can make is \$0.00 if you contribute your \$5.00 and no one else did.

There are many other possibilities, depending on exactly how many people decide to contribute their \$5.00 to the group.

Please make your choice here, by checking one of the following:

_____ I will keep my \$5.00

_____ I contribute my \$5.00 to the group

When you have made your decisions, please turn your packet over and wait for further instructions.

When all 10 participants had made their decisions, the experimenter collected the packets. The experimenter then collected each participant's number and inconspicuously placed the numbers of those who had *not* contributed in an envelope. The rest of the numbers were kept separate.

It was only at this point that the sample from which the data presented here became fixed; it consists of all participants who contributed in Round 1 and were

therefore enabled to punish. Each was randomly assigned to either the *indeterminate* or *determinate* condition. Those in the *indeterminate* condition received the following instructions:

In this round, the choice you make will affect only one other group member. Remember that each group member either did not contribute or did contribute their \$5.00 to the project in the first round. You will at no time learn who contributed and who did not, nor will you learn how many people contributed and how many did not.

To begin this round, each group member who contributed in round 1 will draw a number of another member of the same group, who did *not* contribute. Each of you will know only the number of the person you draw, but will never find out *who* this person is.

You now have the option of punishing this person for not contributing in round 1. Punishment comes at a cost to yourself though. For every \$.20 you pay out, they will be penalized \$1.00 to a maximum punishment of \$5.00 (costing you \$1.00).

Please check off how much you want to punish them.

- don't penalize
- penalize by \$1.00 (cost to you of \$.20)
- penalize by \$2.00 (cost to you of \$.40)
- penalize by \$3.00 (cost to you of \$.60)
- penalize by \$4.00 (cost to you of \$.80)
- penalize by \$5.00 (cost to you of \$1.00)

Participants were instructed to raise their hand once they had made their decision.

The experimenter approached them, one at a time, with the envelope containing numbers of *non-contributors* and the participant then drew the number of the person for whom they could penalize. All numbers were replaced in the envelope so that in sessions in which over half of participants contributed and thus could penalize, there would always be numbers (of non-contributors) to draw.²

In the *determinate* condition, instructions for contributors were identical except that participants drew the number of the person to be penalized *before* making the decision. In both conditions, after making the choice and drawing a number (in one

sequence or the other), participants were asked to rate on a likert scale (from 1-5) the degree of 1) anger, 2) blame, and 3) sympathy they felt for the non-contributing group member whose number they had drawn.³ Each participant who did not contribute in Round 1 was subject to any punishment selected by contributor(s) who drew their number.

Results

Descriptive results

Of the 144 study participants, 55% (n=77) contributed to the group in Round 1. In Round 2, of the 77 participants who contributed and thus could punish, 53.2% levied some punishment on a non-contributor ($M = \$1.79$, $Mdn. = \$1$).

Penalties

Since the dependent variable of ‘penalty’ was censored at \$0, a Tobit regression was utilized (Tobin, 1958). Our major hypothesis, that contributors would apply harsher penalties in the determinate condition than in the indeterminate condition, was supported, $X^2(1, 77) = 4.90$, $p = .03$. The results are detailed in Table 1, and in Figure 1, which presents a frequency distribution of punishment amounts for the two experimental groups. From the last row of the table, it is apparent that the determinateness manipulation affected the magnitude of penalties as well as the tendency to punish. A greater proportion of participants punished a determinate target than an indeterminate target. Although the modal punishment was \$0 for both conditions, the mean and median penalty was greater in the determinate condition.

Emotional Reactions

We also predicted that contributors would react with greater anger and blame toward determinate non-contributors than toward indeterminate non-contributors. ANOVA results regressing penalties on each of these variables revealed significant differences in self-reported anger and blame ($F(1, 75) = 25.79, p < .01$ and $F(1,75) = 8.29, p < .03$) respectively. There were no differences in self reported sympathy ($F(2, 75) = .162, n.s.$).

Mediational analyses

We predicted that the emotional reactions of anger and blame would mediate the effect of determinateness on punitiveness, and tested whether this was the case using Kenny, Kashy, and Bolger's (1998) test of mediation. All of the criteria for full mediation were satisfied. First, as reported above, there was a significant effect of determinateness on penalty. Second, there was a significant relationship between determinateness and a composite measure of anger and blame obtained by averaging the two measures ($F(1, 75) = 10.51, p < .01$). Third, when we regressed penalty on both determinateness and anger/blame, the effect of anger/blame was significant ($\beta = 1.88$), $X^2(1, 77) = 33.34, p < .01$, but the effect of determinateness vanishes almost completely after anger/blame is controlled for ($\beta = .04, n.s.$). These findings provide clear support for the hypothesis that identifiability affects behavior by evoking stronger emotions toward an identified target than toward an unidentified target.

Discussion

The tendency toward more severe punishment for identifiable perpetrators, as demonstrated by our study, has important implications for public policy, and especially for jury decision making and the court system.

Since criminal sentencing inevitably occurs with an identifiable defendant, a juror might feel anger and blame at a level that is *not* experienced by policy makers, when they established the guidelines of appropriate sanctions for particular offenses. This heightened negative reaction toward offenders at the time of trial, coupled with greater sympathy for identifiable victims, might lead to harsher sanctions for actual cases than those set forth by legal guidelines. On the other hand, in actual court cases, factors that elicit sympathy towards perpetrators, such as a difficult childhood or personal difficulties, could also have greater impact **for identifiable perpetrators at trial than for unidentifiable perpetrators considered at the time when policy-makers determine generic sanction levels.**

Identifiability could also explain an effective strategy of politicians-- drawing public attention toward a particular malevolent individual in order to garner support and mobilize aggressive actions against foreign regimes. Just as focusing on an identifiable victim (e.g. the Brady bill) is exploited to win support of policies protecting victims, the emphasis on Saddam Hussein in political speeches and media coverage could serve as a lightning rod--successfully stirring up anger, thereby motivating a desire to right a wrong. Other causes without such a salient identifiable perpetrator may seem less offensive and less in need of opposition.

More generally, this study supports the general conclusion that emotional reactions to other persons, as well as behaviors directed toward them, depend on a variety of non-normative factors. For example, both punitiveness and giving have been found to depend on background mood states (Lerner et al., 1998; Small & Lerner, 2003).

So far the ramifications of identification have only been demonstrated for empathy toward victims and punishment of perpetrators, but given the strength and the consistency of findings in these two areas, it seems likely to subsequent research will demonstrate a far broader range of applications and support the existence of a more general "identifiable other" effect.

References

Averill, J. R. (1983). Studies on anger and aggression. *American Psychologist*, 38, 1145-1160.

Bentacourt, H., & Blair, I. (1992). A cognition (attribution)-emotion model of violence in conflict situations. *Personality and Social Psychology Bulletin*, 18, 343-350.

Epstein, S. (1985). The implications of cognitive-experiential self-theory for research in social psychology and personality. *Journal for the Theory of Social Behavior*, 15, 283-310.

Fehr, E. & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980-994.

Fetherstonhaugh, D., Slovic, P., Johnson, S.M., & Friedrich, J. (1997). Insensitivity to the value of human life: A study of psychophysical numbing". *Journal of Risk and Uncertainty*, 14, 283-300.

Hamilton, D. L. & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, 103(2), 336-355.

Jenni, K. & Loewenstein, G. F. (1997). Explaining the "Identifiable Victim Effect," *Journal of Risk and Uncertainty*, 14, 235-257.

Kenny, D.A., Kashy, D.A., Bolger N. (1998). Data analysis in social psychology. In D.T. Gilbert, S.T. Fiske, and Lindzey, G. (Eds.). *Handbook of social psychology* (4th ed., pp. 233-265). New York: Oxford University Press.

Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, 24(6), 563-574.

Quigley, B. M., & Tedeschi, J. T. (1996). Mediating effects of blame attributions on feelings of anger. *Personality and Social Psychology Bulletin*, 101, 632-652.

Schelling, T. C. (1968). The life you save may be your own. In S.B. Chase (ed.), *Problems in Public Expenditure Analysis*. Washington, DC: The Brookings Institute.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility and blameworthiness*. New York: Springer-Verlag.

Small, D.A. & Lerner, J.S. (2003). Feeling and Deservingness: Emotion effects on welfare policy preferences. Working Paper, Department of Social and Decision Sciences, Carnegie Mellon University.

Small, D. A. & Loewenstein, G. (2003). Helping the victim or helping a victim: Altruism and Identifiability. *Journal of Risk and Uncertainty*, 26(1), 5-16.

Sherman, S. J., Beike, D. R., & Ryalls, K. R. (1999). Dual-processing accounts of inconsistencies in responses to general versus specific cases. In S. Chaiken and Y. Trope (eds.). *Dual-process theories in social psychology*. New York: The Guilford Press.

Solomon, R. C. (1990). *A passion for justice*. Reading, MA: Addison-Wesley.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26 (1), 24-36.

Van Boven, L., Loewenstein, G., Welch, N. and Dunning, D. (2003). *The illusion of courage: Underestimating social-risk aversion in self and others*. Working Paper. Department of Social and Decision Sciences, Carnegie Mellon University.

Weiner, B. (1995). *Judgments of responsibility*. New York: Guilford.

Footnotes

- ¹ Similar procedures have been used in previous studies of punishing behavior (e.g., Fehr & Gächter, 2000).
- ² Contributors replaced drawn numbers due to the inevitable unevenness of the ratio of contributors to non-contributors in many rounds. Therefore, some non-contributors' numbers were drawn and potentially punished more than once and some were never drawn.
- ³ In addition, we asked, "How likely do you think it is that the non-contributing group member will actually receive the penalty that you chose?" in order to assess the believability of the penalty. The mean responses were 3.27 (identifiable) vs. 3.26 (unidentifiable), suggesting that believability was not affected by the experimental manipulation.
- ⁴ As an exploratory measure, non-contributors were given a hypothetical choice about giving back any amount of their experiment payment to a contributor (either indeterminate or determinate). Identifiability of the contributor had no effect on this hypothetical choice. This null result is unsurprising given that participants were likely insufficiently engaged emotionally in the hypothetical task.

Table 1

Condition	Indeterminate Non-contributor (n=38)	Determinate Contributor (n=39)
Mean	\$1.29	\$2.28
Standard Deviation	\$1.92	\$2.21
Median	\$0.00	\$1.00
Mode	\$0	\$0
Percent of \$5.00 (maximum) penalties	15.8%	33.3%
Percent of \$0.00 (minimum) penalties	60.5%	35.9%

Figure 1

