Journal of Behavioral Decision Making
J. Behav. Dec. Making, **18**: 311–318 (2005)
Published online in Wiley InterScience
(www.interscience.wiley.com). **DOI**: 10.1002/bdm.507

The Devil you Know: The Effects of Identifiability on Punishment

DEBORAH A. SMALL[1]* and GEORGE LOEWENSTEIN[2]
[1]*University of Pennsylvania, USA*
[2]*Carnegie Mellon University, USA*

ABSTRACT

Prior research has confirmed Thomas Schelling's observation that people are more sympathetic and hence generous toward specific identified victims than toward ''statistical'' victims who are yet to be identified. In the study presented in this article we demonstrate an equivalent effect for punitiveness. We find that people are more punitive toward identified wrongdoers than toward equivalent, but unidentified, wrongdoers, even when identifying 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 an identified than toward an unidentified target. Consistent with such an account, the effect of identifiability on punishing behavior was mediated by self-reported anger. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS punishment; emotion; identifiability; public goods game

INTRODUCTION

> . . . 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 sympathy differs reliably toward actual 'identified' victims on the one hand, and more abstract 'statistical' victims on the other (Fetherstonhaugh, Slovic, Johnson, & Friedrich, 1997; Kogut & Ritov, 2005a; 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

* Correspondence to: Deborah A. Small, The Wharton School, 700 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340, E-mail: deborahs@wharton.upenn.edu

Contract/grant sponsors: Russell Sage Foundation; Center for Integrated Study of the Human Dimensions of Global Change.

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 article.

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 circumvent the problem that identifying a victim generally means providing information about him or her, so it is always possible that any observed increment in sympathy 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.

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 of the subjects—the ''victims''—were made to return the money. We then gave each of the subjects who had retained the $10 the opportunity to share their money with one of those who had lost their endowment. In the 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 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 identifiable condition as in the unidentifiable condition, despite the fact that identifying the victim provided no information about them. Follow-up research, in which we raised money for a charitable organization, Habitat for Humanity, revealed a similar effect in a more naturalistic setting.

In this article, we focus on punitiveness, rather than altruism, as another form of retributive response that is triggered when an actor has inflicted harm. Fehr and colleagues have examined the role of punishment in regulating economic exchanges (see Fehr & Gachter, 2000; Fehr & Rockenbach, 2003). Their experimental evidence demonstrates that people often are willing to punish ''free-riding'' in social dilemma situations, even when punishment is costly, suggesting motives beyond material self-interest (Fehr & Gachter, 2000). Identifiability and the emotions inspired by it may be one important factor that triggers such motives.

Beyond generalizing the earlier work beyond reactions to victims, the current study also examines whether any observed differences in the punitiveness exhibited toward identified and unidentified perpetrators would be mediated by different reported 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. However, this remains an untested proposition.

Historically, emotions were a taboo topic in economics (see Elster, 1998). However, there has been a recent surge of interest as economists have recognized that many anomalies of the economic model can be explained by emotions. For instance the bargaining literature has emphasized the trade-off between the cognitive reward of monetary relevance and the emotional satisfaction of punishing unacceptable offers (Bosman & van Winden, 2002; Bosman, Sutter, & van Winden, 2005; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Similarly, stock market anomalies and investment decision have been attributed to emotion biasing effects (Hirschleifer & Shumway, 2003). Finally, research on social preferences has emphasized the role of feelings such as ''warm glow'' and envy in preferences for equitable outcomes (Andreoni, 1995; Fehr & Schmidt, 2000). We contend that the representations of specific, identifiable targets are highly affect-laden, engaging the perceiver at a particularly intense level. Such a mediating role of affect, however,

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 emotion to examine in the context of punitiveness, we assumed, would be anger. Psychological research has shown that perceived intentional harm evokes anger (e.g., Bentacourt & Blair, 1992). Anger, like sympathy, is a moral emotion, which can produce strong inferences of blame (Averill, 1983; Weiner, 1995). Moreover, feelings of anger naturally induce a desire to punish (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.

THE 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 individual participants 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 decisions about cooperating and punishing which affected their actual payoffs.

We predicted that people would punish an *identified* non-contributor more severely than an *unidentified* non-contributor. Second, when given a choice to penalize a non-contributor, participants would react with greater anger toward an *identified* non-contributor than toward an *unidentified* non-contributor. Finally, we predicted that the effect of identifiability on punishment would be mediated by feelings of anger.

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 to 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 nine group members must decide whether to contribute your $5 to the group or to keep it for yourself. If you

[1]Similar procedures have been used in previous studies of punishing behavior (e.g., Fehr & Gachter, 2000).

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 ($9 \times \$1.25$).
- 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, buy 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 *unidentified* or *identified* condition. Those in the *unidentified* 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.

In a minute, 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 *identified* 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. Each participant who did not contribute in Round 1 was subject to any punishment selected by contributor(s) who drew their number.[3]

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 ($M = \$1.79$, $Median = \$1$).

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 identifiable condition than in the unidentifiable condition, is supported, $X^2\ (1, 77) = 4.90$, $p = 0.03$. 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 identifiability manipulation affected the magnitude of penalties as well as the tendency to punish. A greater proportion of participants punished an identifiable target than an unidentifiable target. Although the modal punishment was \$0 for both conditions, the mean and median penalty was greater in the identifiable condition.

Emotional reactions

We also predicted that contributors would react with greater anger toward identified non-contributors than toward unidentified non-contributors. Participants did, in fact, report feeling more angry in the identifiable

Table 1. Identifiability of the non-contributor augmented both the tendency and the magnitude of penalties

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

[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. This did not expose us to much risk of ever having to punish someone negatively because it was extremely unlikely that someone would be punished by more than two contributors, and, such doubling up of punishment only occurred to people who did not contribute in a situation in which more than half of the group contributed. In this situation, the minimum earnings of a non-contributor, prior to punishment, was \$11.25. In fact, no participant in our study had zero or negative earnings.

[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) versus 3.26 (unidentifiable), suggesting that believability was not affected by the experimental manipulation.

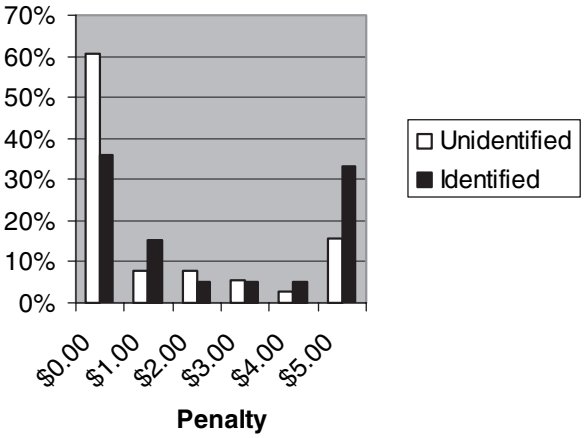

Figure 1. Dark bars represent the proportion of contributors who penalized each possible penalty ranging from $0–5 in the identified condition; light bars represent the proportion that penalized that amount in the unidentified condition

condition ($M = 2.56$, $SD = 1.59$) than in the unidentifiable condition ($M = 1.66$, $SD = 0.71$), $t(75) = -3.59$, $p < 0.01$. Similarly, they reported that they blamed the non-contributor more in the idenitifiable condition ($M = 2.92$, $SD = 1.38$) than in the unidentifiable condition ($M = 2.29$, $SD = 0.98$), $t(75) = -2.31$, $p < 0.03$. There are no differences in self reported sympathy, $t(75) = 0.12$, $p = 0.91$.

We predicted that the emotional reaction of anger would mediate the effect of identifiability on punitiveness. Table 2 presents results regressing penalty on both identifiability and anger. Notice that the effect of anger is significant, but the effect of identifiability vanishes almost completely after anger is controlled for. A Sobel test confirms that the beta weight is significantly reduced ($z = 2.83$, $p < 0.01$) (Sobel, 1982). 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.[4,5]

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.

Table 2. Tobit Regression on Penalties (N = 77)

Dependant variable: Penalties	(1) Baseline	(2) Adds anger
Intercept	−0.51 (0.70)	−3.75* (0.88)
Identifiability	1.96* (0.88)	0.04 (0.71)
Anger	—	1.88* (0.33)

*Significant at the 0.05 level.
Standard errors are reported in parentheses.

[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 unidentifiable or identifiable). 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.
[5]Full mediation was also realized using OLS regression in lieu of Tobit and when a logit regression was performed with a dummy variable for penalty (1 = penalty, 0 = no penalty).

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 lightening 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. In separate work discussing the implications of the identifiable victim and perpetrator effects for public finance (Loewenstein, Small, & Strnad, in press), we show how ''iconic victims and perpetrators'' who attract public attention often provide the impetus for changes in public policy.

The current research extends previous findings about identifiability to a new target and emotional reaction, but it does not address the mechanism by which identifiability amplifies emotional reactions. One possible explanation is that, before the perpetrator's number is drawn, she is perceived as part of a group, rather than an individual case. Kogut and Ritov (2005a, b) have conducted several studies that identified victims by showing a picture of a face in the ''identified'' conditions, and also varied whether the target is a single or a group of victims, and have found that identification of victims only amplifies caring in the case of a single individual. Furthermore, in the identified condition, people give more to a single victim than to a group of victims when evaluated separately (violating dominance), but the pattern is reversed when both the single and group of victims are evaluated jointly, presumably because the dominance relation becomes apparent (Kogut & Ritov, 2005b). They argue that it is necessary to have both identifiability *and* singularity to obtain a heightened affective reaction, and that this reaction is diminished when participants are forced into a deliberative mindset, as in joint evaluation (see also Small, Loewenstein, & Solvic, 2005). Although all of the targets in the current article were single individuals, viewed in the light of Kogut and Ritov's findings, it seems possible that their individuality only becomes manifest after their number has been drawn.

More generally, we believe that identification amplifies feelings and behavior because it makes a situation more concrete and thus reduces the social distance between judges and targets. Social distance may be reduced by several factors, including both identifiability and singularity, but also by similarity and physical proximity.

This study also supports the general conclusion that emotional reactions to other persons, as well as the effects that they elicit, can be influenced by a variety of non-normative factors. For example, punitiveness has been found to depend on randomly induced background mood states that should be irrelevant to the decision at hand (Lerner, Goldberg, & Tetlock, 1998).

So far the ramifications of identification have only been demonstrated for sympathy 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.

ACKNOWLEDGEMENTS

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. We thank Sam Issacharoff (for,

among other things, suggesting the title), Linda Babcock, Margaret Clark, Jennifer Lerner, Uri Simonsohn, and Roberto Weber for their help.

REFERENCES

Andreoni, J. (1995). Warm-Glow versus Cold-Prickle: The effects of positive and negative framing on co-operation in experiments. *Quarterly Journal of Economics, 110*(6), 1–21.

Averill, J. R. (1983). Studies on anger and aggression. *American Psychologist, 3*, 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.

Bosman, R., & van Winden, F. (2002). Emotional hazard in a power to take experiment. *The Economic Journal, 112*, 147–169.

Bosman, R., Sutter, M., & van Winden, F. (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology, 26*, 407–429.

Elster, J. (1996). Rationality and the emotions. *Economic Journal, 106*(438), 1386–97.

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–94.

Fehr, E., & Schmidt, K.(1999). A theory of fairness, competition and cooperation. *The Quarterly Journal of Economics, 114*, 827–868.

Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature, 422*, 137–140.

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.

Hirschleifer, D. A., & Shumway, T. (2003). Good day sunshine: stock returns and the weather. *Journal of Finance, 3* (June), 1009–1032.

Kogut, T., & Ritov, I. (2005a). The ''identified victim'' effect: an identified group or just a single individual? *Journal of Behavioral Decision Making, 18*, 157–167.

Kogut, T., & Ritov, I. (2005b). The singularity effect of identified victims in separate and joint evaluation. *Organizational Behavior and Human Decision Processes, 97*, 106–116.

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.

Loewenstein, G., Small, D. A., & Strnad, J. (in press). Statistical, Identifiable, and Iconic Victims. In J. Slemrod & E. McCaffery (Eds.), *Behavioral Public Finance*.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game, *Science, 300*, 1755–1758.

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.

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.

Small, D. A., Loewenstein, G., & Slovic. (2005). Sympathy and callousness: the impact of deliberative thought on donations to identifiable and statistical victims *Manuscript under review.*

Smith, A. (2000). *The Theory of Moral Sentiments*, New York: Prometheus Books. (Original work published 1723).

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological Methodology 1982* (pp.290–312), San Francisco: Jossey-Bass.

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.

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

Authors' biographies:

Deborah A. Small is an assistant professor of Marketing at the Wharton School of the University of Pennsylvania.

George Loewenstein is a professor of Economics and Psychology at Carnegie Mellon University.

Authors' addresses:

Deborah A. Small, The Wharton School, 760 J. M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340.

George Loewenstein, 208 Porter Hall, Carnegie Mellon University, Pittsburgh, PA 15213.

