Decision

Are Preferences for Allocating Harm Rational?

Alexander L. Davis, John H. Miller, and Sudeep Bhatia

Online First Publication, April 3, 2017. http://dx.doi.org/10.1037/dec0000076

Decision

2325-9965/17/$12.00 http://dx.doi.org/10.1037/dec0000076

Are Preferences for Allocating Harm Rational?

Alexander L. Davis, John H. Miller, and Sudeep Bhatia
Carnegie Mellon University

The allocation of nonmonetary harm is an important—yet understudied—domain of choice. Using a modified Dictator Game, we asked 27 participants to allocate a harmful event (time of putting their hand in ice water) between themselves and an anonymous stranger. We found substantially less coherent, and more egalitarian, preferences compared to other studies that ask participants to allocate monetary endowments. Specifically, 26% of participants made choices inconsistent with utility maximization, and 78% of participants behaved in an egalitarian manner. In comparable studies of monetary gains, only 2% were inconsistent and 30% egalitarian. The results suggest that the focus on monetary gains likely overestimates the rationality of other-regarding preferences and underestimates egalitarianism.

Keywords: altruism, prosocial behavior, rationality, revealed preference

Research in the last two decades has illuminated our understanding of costly, other-regarding behavior (Andreoni & Vesterlund, 2001; Benabou & Tirole, 2006; Fehr & Schmidt, 2006; Fisman, Kariv, & Markovits, 2007; Forsythe, Horowitz, Savin, & Sefton, 1994; Frey & Bohnet, 1995). Most of this work has used experiments like the Dictator Game to show that, in general, individuals are willing to share a monetary windfall with an anonymous stranger in a way that is consistent with some set of well-behaved (utility-maximizing) preferences (Andreoni & Miller, 2002). This prior work focused almost exclusively on the allocation of monetary gains. Here we examine how individuals allocate nonmonetary bads, examples of which range from the heroic, such as an individual going off to war, to the more mundane, like taking out the trash.

We emphasize nonmonetary bads for a variety of reasons. First, as discussed above, it is a domain where decisions are frequently made, including examples that can be quite common (like chores), to rather astounding, such as when an individual risks his life to save the lives of strangers (see, for just one example, Dovidio, Piliavin, Schroeder, and Penner's (2006) discussion of Paul Rusesabagina who helped Tutsis and Hutus escape the Rwandan genocide by letting them stay in his hotel). Between these extremes include the donation of blood and organs, serving in the armed forces, participating in a revolution, or even becoming a suicide bomber. Notwithstanding the importance and real-life consequences of nonmonetary bads, little systematic study has been devoted to understanding behavior in this domain. Second, there are rarely markets or prices for these bads, allowing us to test whether existing models of other-regarding preferences generalize to a novel context. Third, decisions about allocating harm likely invoke types of reasoning (e.g., moral-deontological) that may be inconsistent with utility maximization. Finally, previous work in psychology has found that people are willing to experience harm in place of another person. For example, a study by Batson et al. (1983) found that 65% of participants volunteered to take electric shocks in place of another participant (who was actually an experimental

Alexander L. Davis, Department of Engineering and Public Policy, Carnegie Mellon University; John H. Miller and Sudeep Bhatia, Department of Social and Decision Sciences, Carnegie Mellon University.

Sudeep Bhatia is now at the Department of Psychology, University of Pennsylvania.

We would like to thank Terence Einhorn, John Sperger, and April Jintao for their valuable assistance in helping conduct the research.

Correspondence concerning this article should be addressed to Alexander L. Davis, Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: alexander.l.davis1@gmail.com

confederate), and on average agreed to take more than 50% of the shocks (although none of the shocks were actually implemented). In comparison with the typically more modest rates of generosity observed in Dictator Games using monetary gains, these proportions raise the possibility of greater generosity for allocations of nonmonetary harm. However, differences in experimental design, including the use of deception, lack of anonymity, and available information about the victim, limit comparability across these studies. In this work we provide a more direct comparison using a Dictator Game with allocations of nonmonetary harm.

Our research strategy uses an approach created by Andreoni and Miller (2002), where participants are asked to make a series of choices that involve allocating a resource between themselves and another person. In our modification, each choice involves a different tradeoff between harming another person and oneself. Given these choices, we test whether the participant's behavior is consistent with the *Generalized Axiom of Revealed Preference* (GARP; Afriat, 1967; Varian, 1982). GARP captures the intuition that inconsistency in observed choices can reveal irrational decision making. For example, if a decision-maker chooses a bundle of goods *A* when an alternative bundle *B* could have been chosen at the same price, the choice reveals that *A* is preferred to *B* (or indifferent to *B*). If we then observe the same decision-maker choosing bundle *B* when *A* is cheaper than *B*, an inconsistency would arise. Remarkably, a formalization of this simple idea captures all of the observable consequences of the standard utility theory used by economics.

In addition to examining the consistency of choices, we also explore the diversity of preferences that arise in this context. Specifically, Andreoni and Miller (2002) found that participants' preferences are well captured by three types of functions: Selfish, Leontief, and Perfect Substitutes. If we call the dictator's utility function $U(x_s, x_o)$, where x_s is the amount of good allocated to the dictator, and x_o is the amount of good allocated to the other person, then a Selfish participant has the simple function $U(x_s, x_o) = x_s$, which is only dependent on the payoff to herself. The utility function for a dictator with Leontief preferences is $U(x_s, x_o) = \min(x_s, x_o)$; that is, the utility experienced is only as large as the minimum amount of good allocated to the pair. As a result, a participant with Leontief preferences tries to equalize the outcomes for herself and the other participant, behaving in an egalitarian manner. Finally, the utility function for a dictator with Perfect Substitute preferences is $U(x_s, x_o) = x_s + x_o$, meaning the outcomes for herself and the other person are treated interchangeably. A participant with Perfect Substitutes preferences will maximize the total amount of the good obtained by the pair.

Method

Participants

In advance we aimed to collect 25 pairs of dictators and receivers, a number we considered large enough to detect interesting effects given our budget constraints. Twenty-seven pairs of participants were ultimately recruited through the Center for Behavioral Decision Research website. Sessions were run in groups of four to eight participants (two to four pairs). The average age of the dictators was 24 years (with a range from 18–59), 14 were women.[1]

Procedure

Participants were randomly assigned an identification number unknowable to others using shuffled opaque sealed envelopes. After random assignment, participants were told that they would be paid $10 in cash at the end of the experiment. They were informed that they would be split into two rooms (based on whether they had an even or odd identification number), and that each odd-numbered participant (dictator) would make a decision that would affect an even-numbered participant (receiver), but that even-numbered participants would not make any choices that affected the odd-numbered participants.[2] All participants were assured that their choices throughout the experiment would be anonymous. Before making their choices, participants came to the "Ice

[1] All de-identified data, R code for analyses, and materials associated with the study are available through the Open Science Framework at http://osf.io/82jpk.

[2] We note that even though the receiver was anonymous, there was exactly one receiver, implying the existence of an identifiable victim effect. However, such an effect is also present in other Dictator Game studies.

Table 1
Parameters for the Eight Choices Made by Each Dictator, and Choice Patterns of Time Not in Water (x_s, x_o) for Each of the Preference Types

Decision	Tokens (M)	Hold (m_s)	Pass (m_o)	$\acute{M}$	p	Selfish	Leontief	PerfectSub
1	60	1	1	60	1	(60, 0)	(30, 30)	(0–60, 0–60)
2	60	1.5	1	60	1.5	(60, 0)	(24, 24)	(60, 0)
3	60	2	1	60	2	(60, 0)	(20, 20)	(60, 0)
4	60	3	1	60	3	(60, 0)	(15, 15)	(60, 0)
5	60	1	1.5	40	.67	(40, 0)	(24, 24)	(0, 60)
6	60	1	2	30	.50	(30, 0)	(20, 20)	(0, 30)
7	60	1	3	20	.33	(20, 0)	(15, 15)	(0, 20)
8	80	1	1	40	1	(40, 0)	(20, 20)	(0–60, 0–60)

Water Experience" section of the questionnaire, where they were asked to raise their hand to notify the experimenter. The experimenter then walked to the participant with an insulated container of ice-water maintained at 3–5 °C, and had the participant submerge her hand in the water for 5 s timed with a digital stopwatch that was observable by the participant.

Dictators then made eight allocation decisions for their single paired receiver.[3] All participants then completed an auction to determine how much they valued avoiding the ice-water experience. When all participants were finished, a random number generator picked one of the eight decisions to be implemented. Once this procedure was finished, the receivers were called into the payment room, one at a time, and received their payoffs, including submerging their hand in ice-water. The participant's auction decision was then checked to see if she had won the auction, and if so, her hand was submerged in ice water for 60 s and she was paid accordingly. Finally, they were paid their $10 show-up fee. This process was repeated for dictators, who put their hand in the ice-water for the duration of the time specified by their randomly selected choice and then received the outcome of the auction. Additional details on the experimental procedure are presented in the appendix.

Choice Task

The main task was a series of eight decisions, randomly ordered across subjects. Each decision required the dictator to make an allocation of a budget of either 60 or 80 tokens between herself and her assigned, anonymous receiver. Each dictator's allocation of tokens was transformed into a period of time she and the receiver would have to submerge their hands in ice water if that particular decision was the one randomly chosen to be executed at the end of the experiment. For each of the eight choices, each token was assigned a unique combination of hold and pass values. The hold value was the amount of time (in seconds) that the dictator would have to keep her hand in ice water for each token so allocated, while the pass value was the amount of time (in seconds) that the receiver would have to keep his hand in ice water for each token allocated to him. The eight decisions are summarized in Table 1. For example, one choice asked dictators to complete the following task:

> Divide 60 tokens: Hold _____ tokens at 1 second(s) each, and Pass _____ tokens at 1 second(s) each.

The choices were made via a computer interface driven by the Qualtrics system (see Figure 1 for a sample screen). Dictators were required to press a "validate" button before submitting their choices. Validation calculated the actual submersion times implied by the choices and displayed these to the dictator to help her understand the full consequences of her decision. Validation also verified that the choices met three constraints that were known by the dicta-

[3] Receivers also made predictions about the average behavior of dictators, but we do not report predictions here to keep our discussion focused on the behavior of the dictators. Data are available from the Open Science Framework.

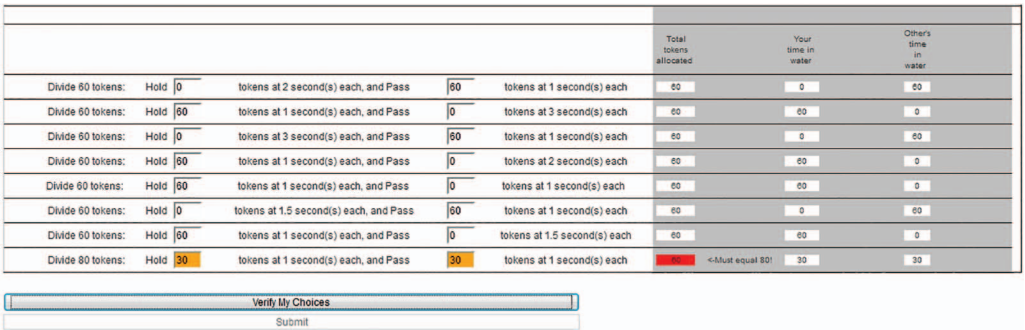

Figure 1. Sample screen of the eight choices presented to the dictators (each dictator received a randomly ordered table), with a constraint violation in the last choice. See the online article for the color version of this figure.

tors in advance: (a) that the total tokens allocated in any given choice exactly equaled the total number of tokens available in that choice,[4] (b) that allocations were nonnegative, integer amounts, and (c) that the allocations would not have any participant receiving more than 60 s of immersion time in the ice water, a constraint imposed to avoid permanent harm. Participants were automatically notified with error messages and colored textboxes if they violated any of these constraints. If all of the allocations over the eight decisions were valid, a button became enabled that allowed the dictators to submit their decisions, though they could still alter (and revalidate) their choices at that time before submission.

Underlying Choice Design

Each dictator's choice task was to divide M tokens between self and other, with multipliers m_s (hold value) and m_o (pass value). The multipliers determined the final amount of time that the subject would have to endure the unpleasant task, such that if K tokens were kept, the dictator would immerse her hand in ice water for $y_s = K \times m_s$ seconds and the receiver (other) for $y_o = (M - K) \times m_o$ seconds. For example, suppose the dictator was asked to divide 60 tokens ($M = 60$), where each token was worth 2 s to oneself ($m_s = 2$) and 1 s to the other person ($m_o = 1$). If the dictator keeps 20 tokens ($K = 20$), then the dictator gets 40 s ($= 20 \times 2$) and the receiver gets 40 s ($= 40 \times 1$).

Standard GARP analysis and the estimation of utility functions requires "positive" choice objects, that is, objects for which utility is monotonically increasing. Thus, we transform the outcome variable in the experiment—time in water—so that it represents a positive good—time *not* in water. Let x_s represent time for self *not* in water and x_o time for other *not* in water. Therefore $x_s = 60 - y_s$ and $x_o = 60 - y_o$ because the maximum amount of time in ice water for either person is bounded above at 60 s. Furthermore, $p = \frac{m_s}{m_o}$ is the normalized price of time not in water (price of giving) and $\acute{M} = 60 + 60 \times \frac{m_s}{m_o} - m_s \times M$ is the normalized income, or how much time not in water the dictator can allocate to herself. Figure 2 shows all of the transformed budget sets across the eight choices used in the experiment. The eight choices were chosen so that the multiple budget-set overlaps would provide many opportunities for potential GARP violations. More information about the transformation, transformed budget set, and a discussion of WARP/GARP violations are provided in the appendix.

[4] GARP allows choices on full budget set versus only the budget constraint. To allow this here, dictators would need to be given extra tokens (implying additional immersion time) that they could allocate to either themselves or the receiver. Given the extra complication of including this option, and our sense that it would not be exercised given choices observed in related studies (see Fisman et al., 2007 and Andreoni & Miller, 2002), we constrained choices as per the text. All of the dictators gave positive valuations for their Willingness to Accept additional harm based on the auction, and this is consistent with the notion that the option for immersion time beyond the minimum required would not have been utilized.

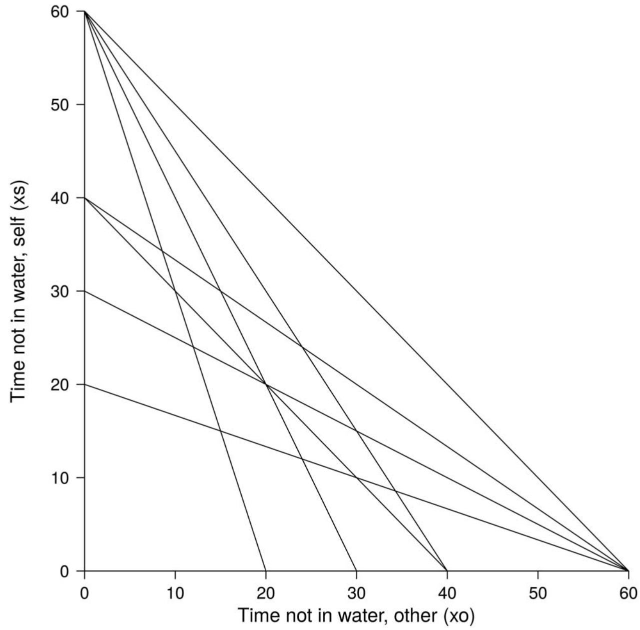

Figure 2. Budget sets of the eight choices transformed to the space of time in not water for self (x_s) and other (x_o).

Auction

At the end of the experiment, participants completed a second-price auction in which the lowest bidder in a session would receive a payment (determined by the amount bid by the second-lowest bidder to encourage truthful revelation) to immerse her hand in ice water for 60 s. We use the bids in this auction to determine each dictator's minimum Willingness to Accept (WTA) for the ice-water-immersion experience, and to control for possible heterogeneity in disutility from this experience.

Results

Our results focus on three main questions: (a) what choices do dictators make in the domain of harm, (b) are these choices rationalizable (consistent with GARP), and (c) what do the implied preferences look like? To begin the analysis, we provide a brief summary of the choices made by dictators, and from there we explore the question of rationalizability. Finally, we take a more in-depth look at the types of preferences that we observe across the dictators.

Choice Behavior

First we consider the choices made when the prices of giving time not in water were both equal to one, as in the standard Dictator Game. In our experiment, dictators had to make two such allocations, one with 60 and one with 80 tokens. With 60 tokens, the dictators keep a mean of 24.4s of putting their hand in ice water, thus, taking on 41% of the bad (time in water; 95% confidence interval, CI [32% to 49%]),[5] or equivalently, 59% of the good (time not in water; 95% CI [51% to 68%]). With 80 tokens, they keep a mean of 34.1 s or 43% of the bad (95% CI [38% to 47%]; given the constraint preventing allocations of more than 60 s, at the very least the dictator had to keep 25%). Because the good is defined in terms of time relative to 40 s of time not in water (i.e., $\acute{M} = 40$), this implies that the dictators took a mean of 65% of the good in this latter case (95% CI

[5] Because the distributions of sharing were skewed, 95% confidence intervals are calculated using a simple nonparametric bootstrap.

[57% to 73%]).[6] The generosity observed here is greater than observed in monetary good experiments, for example, in Andreoni and Miller (2002), participants keep around 76% and 83% of the good (money) under similar conditions. For further comparison, in regular Dictator Games, Forsythe et al. (1994) find that participants keep an average of around 77–78% of the monetary endowment and Camerer (2003; chapter 2) reports an average of 80%. Participants in our study kept about 60% of the good, a result that is substantially different from what is observed with monetary gains.

Of the 27 dictators, 16 of them allocated the identical percentage of the bad under both of the above token levels (with 15 of them keeping 50%, and one keeping 25% of the time in water). Two of the dictators (7%) took on more of the bad than they gave in at least one of these two choices, behavior that is almost never observed in the case of the standard Dictator Game. Although 22 (81%) of the dictators made at least one choice where they gave the receiver more harm than they kept for themselves, only 7 (26%) dictators did this across all of their choices. We find that in 9% of all the total possible choices the dictators keep more of the bad than they give, in 47% of the choices they give more of the bad than they keep, and in the remaining 44% of choices the bad was equally distributed. Again, these results support the idea that participants were much more generous in the context of nonmonetary harm than monetary gain.

All of the dictators had a positive WTA for 60 additional seconds in water, as determined by the second-price auction, indicating they disliked the experience. The lowest WTA was \$0.50 and the highest, an extreme outlier, was \$1,000,000 (the second highest was \$10). Removing the outlier, the mean WTA was \$3.23 and the median was \$2.75. WTA was uncorrelated (Pearson's $r = -.03$) with the mean of the income-normalized amount of time not in water $\left(\frac{x_s}{\acute{M}}\right)$ kept by the dictator, suggesting that differences in altruism across dictators cannot be explained easily by their sensitivity to, or displeasure from, the ice-water task.

Rationalizability

If choices satisfy GARP, then they are rationalizable in the sense that they could have resulted from the maximization of utility on a set of well-behaved preferences (Afriat, 1967; Varian, 1982). GARP violations occur when a subject, directly or indirectly, alters her preferences between bundles simultaneously available under different budget sets. To benchmark the likelihood of GARP violations, we generate random choices on the budget sets we used in our experiment and check for GARP (Bronars, 1987). Using this procedure, we find that on average 93% (N = 1,000) of random choices result in at least one GARP violation. In our experiment, 48% (13/27) of the dictators had at least one GARP violation.

Afriat (1972) developed the Critical Cost Efficiency Index (CCEI) as a measure of the degree of GARP violation. The CCEI for any choice is the proportionate reduction in spending needed to purchase a revealed-preferred bundle, that is, how much less could the subject spend to receive a bundle that is at least as good as what she bought. Thus, if the CCEI is 1.0, no reduction is possible, but as the CCEI falls, the GARP violation becomes more severe. Following Varian (1994), we associate an individual's lowest CCEI as a measure of the degree of GARP violation. Slightly over half of our subjects showed no violations of GARP. A CCEI of 0.95 has been used in the literature to mark significant GARP violations. Based on this threshold, we find that 26% (7/27) of our subjects had behavior that cannot be rationalized as arising from utility-maximization on well-behaved preferences. By contrast, Andreoni and Miller (2002) explored monetary gains and found that less than 2% (3/176) of their participants had such violations. Additional comparisons with prior work are provided in the appendix.

Preferences

In the prior work of Andreoni and Miller (2002), a large amount of heterogeneity was observed among the apparent preference types of the subjects. One simple way to classify preferences is to consider three prototypical forms of behavior: (a) Leontief preferences, where the dictator seeks to allocate the bad so that each person gets an equal amount, (b) Selfish preferences, where the

[6] For example, if the dictator kept 20 tokens and gave 60 with $\acute{M}$ = 40, then she took all of the good ($1 = \frac{60-20}{40}$).

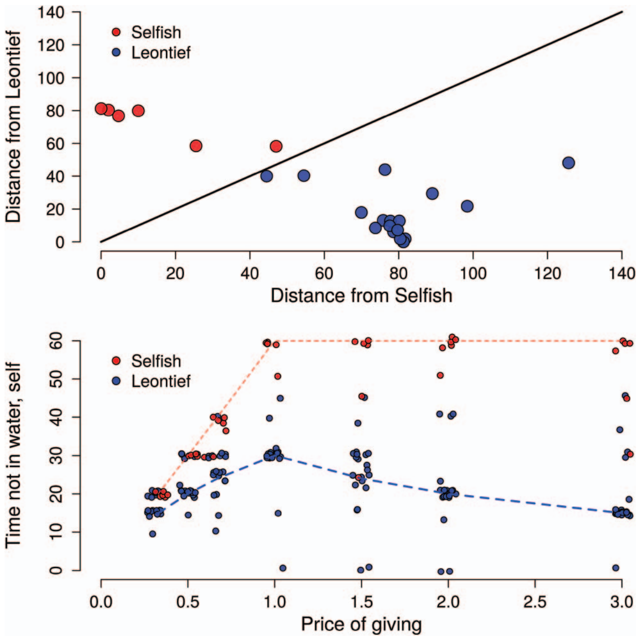

Figure 3. Top panel: Euclidian distance of each dictator's choices to prototypical pure Leontief and Selfish choices. Bottom panel: Dictators' decisions to keep time not in water as a function of the price of giving time not in water for decisions 1–7 (with 60 tokens). The dashed red and blue lines indicate the behavior predicted by purely Selfish or Leontief preferences (respectively). See the online article for the color version of this figure.

dictator gives the other subject the maximum amount of time in water possible, and (c) Perfect Substitutes where the dictator attempts to minimize the aggregate immersion time regardless of which person is subjected to the harm. Using these three prototypes, we can classify dictators as either Strong, that is, the observed behavior is fully consistent with the particular preference prototype, or Weak, that is, the observed behavior is best captured by the prototype in terms of mean Euclidean distance (see Figure 3). Using this metric, Table 2 classifies the observed behavior of the 27 subjects.[7]

Table 2
Distribution of the 27 Dictators by Closest Match to Prototypical Preference Forms

	Leontief	Selfish	Perfect substitutes
Strong	4 (15%)	1 (4%)	0 (0%)
Weak	17 (63%)	5 (19%)	0 (0%)
Total	21 (78%)	6 (22%)	0 (0%)

The distribution of apparent preference types is quite different from what has been previously observed in monetary-gain experiments. Here, 22% of the dictators are Selfish, whereas under monetary-gains in Andreoni and Miller (2002) about half were selfish; 78% are Leontief, versus only 30% under monetary-gains; and, finally, we observed no one behaving as if they had Perfect Substitutes, whereas 22% of subjects followed this prototype under monetary-gains.[8] In the realm of nonmonetary bads, a

[7] If we admit the possibility of pure Leontief preferences, but with distribution ratios (for time not in water) other than 1/1 (self/other), our overall classification changes slightly. In this case, counting both weak and strong types, we find 5 Selfish, 13 Leontief at 1/1, 6 Leontief at 3/2, and 3 Leontief at 3/1.

[8] Fisman et al. (2007) looked at only strong types, and found 20% of their participants being Selfish, 2.6% Leontief, and 2.6% having Perfect Substitutes. Andreoni and Vesterlund (2001) find strong types at 21.2% Selfish, 16.2% Leontief, and 0.07% Perfect Substitutes, and the total distribution of types at 44%, 35%, and 21%, respectively.

large majority (78%) of dictators make choices consistent with sharing the harm equally.[9]

Discussion

Individuals often make choices about how to allocate bad events, sometimes in the mundane activities of daily household life, and at other times over more substantial events like serving in the armed forces or rescuing someone in danger. Notwithstanding both the prevalence and importance of such decisions, they have received a surprising lack of attention in the experimental literature. Here, we focus on nonmonetary choices over bad events. We find that subjects are much more generous in such choices—about three-quarters of them deciding to share the harm about equally—and that the behavior of about a quarter of the subjects cannot be reconciled as being driven by the maximization of a well-behaved utility function—versus the almost universal ability to do so in the context of monetary gains.

A few psychologists have performed experiments on the allocation of harm in very different domains to better understand the relationship between empathy and altruism (Batson, 1987; Batson et al., 1988; Schaller & Cialdini, 1988), and have also found that individuals tend to share the harm. These experiments were not designed to tease out preference types or violations of rationality, and relied on very different methodologies (including the use of deception), though the coincidence of results across radically different methodologies suggests that such behavior may be relatively robust.

In our experiment, we find preferences that are quite different than previous results using the standard Dictator Game with monetary gains. Of course, our context differs from this prior work on two dimensions: we look at nonmonetary versus monetary and at losses versus gains. There is some evidence (Vohs, Mead, & Goode, 2006) that other-regarding preferences in the context of money tend to be more selfish. The evidence about losses versus gains is less clear. Given these observations, we cannot easily disentangle whether the behavior we are observing here is tied purely to the nonmonetary nature of the choice or the imposition of harm on another. While teasing out these two effects is an interesting question, exploring the joint effect is still of interest given that individuals must often make such choices.

Our results present an interesting theoretical challenge. At the aggregate level, we know that some individuals must be harboring very different sets of preferences over allocating monetary gains versus nonmonetary harm. It is not clear that the existing literature on prosocial preferences (Fehr & Schmidt, 2006) can easily accommodate such observations in a unified framework, because we would need to find a single, metapreference function that displays such switching in these two contexts. Moreover, we observed a breakdown of well-behaved preference maximization in over a quarter of our subjects in the context of nonmonetary harm, and thus, even if we can solve the first challenge, we may still have difficulty explaining the behavior of one out of four subjects.

It may be the case that the imposition of harm on another causes a breakdown of our usual models, at least in some participants. There is some evidence that people prefer indirectly over directly harming others (Royzman & Baron, 2002). That being said, even a simple rule like "do no harm," while perhaps being able to explain the asymmetry of preferences in the monetary-gain versus nonmonetary-loss contexts, would still lead to well-behaved preference maximization within each context. There is also evidence that moral motives, like justice and harm, hold across cultures (Haidt & Graham, 2007), so perhaps the interplay of moral factors might provide some insight here. Alternatively, perhaps monetary-gains versus nonmonetary-losses imply very different cognitive processes, resulting in both asymmetries of preferences and the potential breakdown of optimized choice. For example, moral judgments tend to be the consequence of emotional responses, especially when directly

[9] The preferences across allocating harm to self and other are tied to preferences for harm to self. While we have data on each subject's Willingness to Accept for 60 s of cold-water immersion, we do not have enough additional information to derive an individual's demand curve for the cold-water task. In general, subjects tend not to recognize how much they will adapt to harm (Ubel, Loewenstein, & Jepson, 2005), so we suspect that subjects will not discount additional time as much as they should. Even if they discounted such time radically, the equal sharing of time (vs. either taking it or giving it all) is unexpected.

harming others is involved (Greene & Haidt, 2002), and there are examples of decreased rationality when high-intensity emotions are involved in decision making (Loewenstein & Lerner, 2003).

An alternative explanation involves demand characteristics, where the researcher unknowingly provides subtle cues about how participants are expected to respond (Bardsley, 2008; Levitt & List, 2007; List, 2007; Orne, 1962; Zizzo, 2010). For example, Bardsley (2008) found that dictators are less likely to give to the receiver when taking from the receiver is possible. He argues that when one can only give to others, dictators see their task as being about giving, but when one can take from the receiver, the expected behavior (demand characteristic) is less clear and thus, giving is reduced. Indeed, any participant in our experiment could have chosen to leave the experiment with pay and receive none of the pain from placing her hand in ice water or, assuming the receiver also withdrew, guilt from forcing another person to do so, but no one did. Their right to do this was explained in the informed consent document at the beginning of the experiment. Thus, even the mere participation in our experiment provides a puzzle not easily resolved by standard social-preference theory.

Other-regarding behavior, across a wide variety of contexts, is an important element of our world. Prior work has had a very useful, though near-exclusive, focus on one aspect of this domain, namely monetary gains. Here we extend the exploration of other-regarding behavior into an analysis of the allocation of harmful, nonmonetary events. We find that individuals behave very differently in this new domain, both in terms of the preferences that they reveal and their ability to behave in a rationalizable way.

References

Afriat, S. (1967). The construction of utility functions from expenditure data. *International Economic Review, 8,* 67–77.

Afriat, S. (1972). Efficiency estimates of production functions. *International Economic Review, 13,* 568–598.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica, 70,* 737–753.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics, 116,* 293–312.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics, 11,* 122–133.

Batson, C. (1987). Prosocial motivation: Is it ever truly altruistic? In L. Berkowitz (Eds.), *Advances in experimental social psychology* (pp. 65–122). Cambridge, MA: Academic Press.

Batson, C. D., Dyck, J. L., Brandt, J. R., Batson, J. G., Powell, A. L., McMaster, M. R., & Griffitt, C. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of Personality and Social Psychology, 55,* 52–77.

Batson, C., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology, 45,* 706–718.

Benabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review, 96,* 1652–1678.

Bronars, S. G. (1987). The power of nonparametric tests of preference maximization. *Econometrica, 55,* 693–698.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Dobell, A. R. (1965). A comment on AYC Koo's "An Empirical Test of Revealed Preference Theory." *Econometrica, 33,* 451–455.

Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. (2006). *The social psychology of prosocial behavior*. Mahwah, NJ: Erlbaum Publishers.

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism: Experimental evidence and new theories. In S. Kolm & J. Ythier (Eds.), *Handbook on the economics of giving, reciprocity and altruism* (pp. 615–691). Amsterdam, the Netherlands: Elsevier.

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *The American Economic Review, 97,* 1858–1876.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6,* 347–369.

Frey, B. S., & Bohnet, I. (1995). Institutions affect fairness: Experimental investigations. *Journal of Institutional and Theoretical Economics, 151,* 286–303.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6,* 517–523.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions

that liberals may not recognize. *Social Justice Research, 20,* 98–116.
Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives, 21,* 153–174.
List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy, 115,* 482–493.
Loewenstein, G., & Lerner, J. (2003). The role of affect in decision making. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.), *Handbook of affective science* (pp. 619–642). Oxford, United Kingdom: Oxford University Press.
Orne, M. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.
Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research, 15,* 165–184.
Schaller, M., & Cialdini, R. B. (1988). The economics of empathic helping: Support for a mood management motive. *Journal of Experimental Social Psychology, 24,* 163–181.
Ubel, P. A., Loewenstein, G., & Jepson, C. (2005). Disability and sunshine: Can hedonic predictions be improved by drawing attention to focusing illusions or emotional adaptation? *Journal of Experimental Psychology: Applied, 11,* 111–123.
Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica, 50,* 945–973.
Varian, H. R. (1994). *Goodness-of-fit for revealed preference tests*. University of Michigan CREST Working Paper 13.
Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science, 314,* 1154–1156.
Warshall, S. (1962). A theorem on Boolean matrices. *Journal of the ACM, 9,* 11–12.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13,* 75–98.

Appendix A

Budget Set Transformation

Let x_s represent time for self *not* in water and x_o time for other *not* in water. Therefore $x_s = 60 - y_s$ and $x_o = 60 - y_o$ because the maximum amount of time in ice water for either person is bounded above at 60 s. Thus, the point $(x_s, x_o) = (0, 0)$ represents 60 s of time in water for self and other, and $(x_s, x_o) = (60, 60)$ represents 0 s of time in water for self and other. The set of feasible choices (*budget set*), are defined by the number of tokens and the hold and pass values. Specifically, the budget set includes the possible pairs of time for self not in water (x_s) and time for other not in water (x_o) such that the total amount of time in water allocated matches the number of tokens (weighted by the hold and pass values), and that the time not in water allocated to oneself and the other person are both greater than or equal to zero (i.e., nobody can experience more than 60 s of ice water). Mathematically, a budget set is defined in this context as:

$$\mathbf{B} = \{\{x_s, x_o\} \,|\, \{x_s + x_o \times p = \acute{M}\} \cap \{x_s, x_o\} \,|\, x_s \geq 0 \ \ \& \ \ x_o \geq 0\} \tag{1}$$

where $p = \frac{m_s}{m_o}$ is the normalized price of time not in water and

(*Appendices continue*)

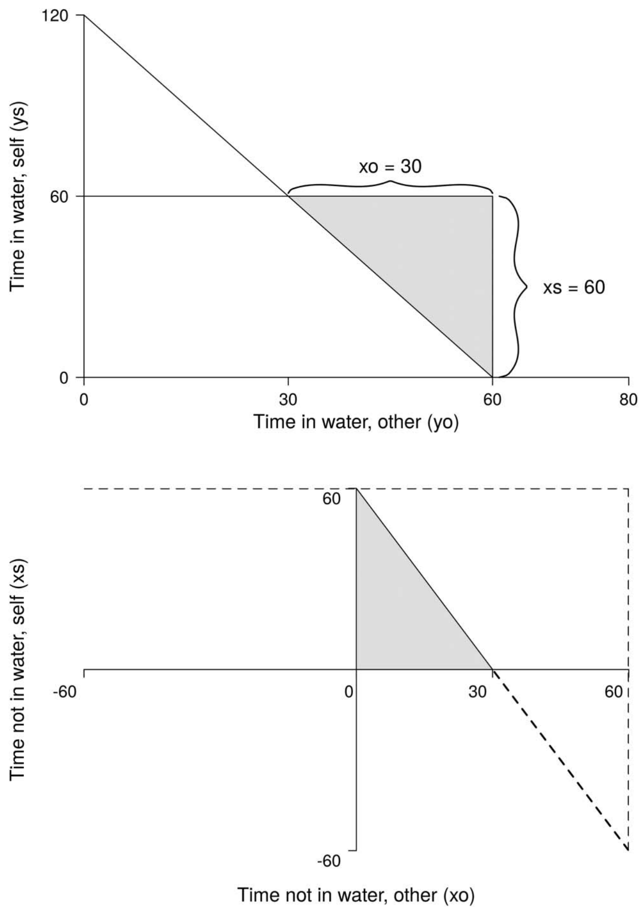

Figure A1. A sample budget set in terms of time in water for self and other (upper), and its equivalent transformation for time not in water (lower). The budget set is $M = 60$, $m_s = 2$, and $m_o = 1$, with the shaded area corresponding to the choices possible in the budget set.

$$\acute{M} = 60 + 60 \times \frac{m_s}{m_o} - m_s \times M \qquad (2)$$

is the normalized income.[10] $\acute{M}$ can be interpreted as the amount of time not in water that the dictator has available to allocate to herself. For example, suppose $m_s = m_o = 1$ and $M = 60$, then the dictator is free to allocate all 60 s of time not in water ($\acute{M} = 60 + 60 \times \frac{1}{1} - 1 \times 60$) in any way desired. For the budget sets we use, $\acute{M} = 60$ when $M = 60$ and $m_s \geq m_o$. On the other hand, when $m_o > m_s$, then the dictator is not free to allocate all of the time not in water to herself. For example, when $m_0 = 3$ and $m_s = 1$,

[10] $\acute{M}$ is how much time not in water the dictator can allocate to herself. This comes from the constraint that, in that situation, the maximum amount of time in water that can be allocated to the other person is 60 s. This gives $y_o = (M - K)m_o = 60$, and as a result, $K = M - \frac{60}{m_o}$. Because $y_s = Km_s$, we get $y_s = \left(M - \frac{60}{m_o}\right)m_s$, or the amount of bad that the dictator can allocate to herself. The amount of good is then $x_s = 60 - y_s = 60 + 60 \times \frac{m_s}{m_o} - m_s \times M$.

(*Appendices continue*)

then the dictator is only free to allocate 20 s of time not in water to herself ($\acute{M} = 60 + 60 \times \frac{1}{3} - 1 \times 60$). This is because each token given to the other participant incurs 3 s of time. This constraint is necessary to ensure that no participant is exposed to more than 60 s of time, to avoid injury.

The upper diagram in Figure A1 represents the nontransformed space of possible choices for a sample budget set $M = 60$, $m_s = 2$, and $m_o = 1$, with the shaded area corresponding to the choices possible in the budget set. The lower diagram in that figure represents the equivalent transformation of this budget set over positive goods. In the lower diagram, the normalized price of giving time not in water to the other person, p, is equal to 2 ($= \frac{m_s}{m_o} = \frac{2}{1}$), and the shaded area corresponds to the transformed budget set over the two positive goods, namely, time not in water for self and other. Rotating the lower diagram around the intersection of the two dotted lines by 180 degrees and using the dotted lines as axes with the transformed origin, shows how the nontransformed space is contained within the transformed one.

Appendix B

Checking WARP

Formally, GARP relies on three notions of revealed preference (Varian, 1982). Consider allocations of time not in water to self and other (x_s, x_o) with the price of x_s normalized to 1 and the price of x_o equal to $p = \frac{m_s}{m_o}$. The total cost of purchasing this bundle is $C = x_s + px_o$. If we observe a decision-maker choose $x^1 = (x_s^1, x_o^1)$ over $x^2 = (x_s^2, x_o^2)$, then it must be the case that $C_1 = x_s^1 + px_o^1 \geq C_2 = x_s^2 + px_o^2$. When this occurs we say that x^1 is *directly revealed preferred* to x^2, because the decision-maker is willing to pay at least as much to get x^1 as she is willing to pay to get x^2. If the inequality is strict, such that $C_1 > C_2$, then x^1 is *strictly directly revealed preferred* to x^2. Finally, if x^1 is directly revealed preferred to x^2, and x^2 is directly revealed preferred to x^3, then x^1 is *indirectly revealed preferred* to x^3. The *Weak Axiom of Revealed Preference* (WARP) states that if x^1 is directly revealed preferred to x^2, then x^2 is not directly revealed preferred to x^1. The *Strong Axiom of Revealed Preference* (SARP) states that if x^1 is revealed preferred (directly or indirectly) to x^2, then x^2 is not directly revealed preferred to x^1. Finally, GARP states that if x^1 is revealed preferred to x^2 (directly or indirectly), then x^2 is not strictly directly revealed preferred to x^1. SARP and GARP both capture the familiar concept of transitive preferences, but only GARP captures the necessary and sufficient conditions for choices to be consistent with utility maximization.

To check WARP, consider the budget set $B = x_s + x_o p \leq \acute{M}$, where x_s is the amount of good (time not in water) kept for the dictator, x_o is the amount of time not in water given to the receiver, $p = \frac{m_s}{m_o}$ is the ratio of hold and pass values of the tokens, and $\acute{M}$ is the normalized income. B can be written as the area beneath the line $x_s = \acute{M} - x_o p$, with intercept $\acute{M}$ and slope $-p$. All points on this line or between it and (0, 0) are within the budget set B. Thus, any point on the budget constraint $x_s = \acute{M} - x_o p$ must be preferred to points $x_s < \acute{M} - x_o p$, as the latter provide strictly less quantities of goods x_s and x_o than the former. Suppose the point $x^1 = (x_s^1, x_o^1)$ lies on the budget constraint of B, such that $x_s^1 + x_o^1 p = \acute{M}$. Then, the point $x^2 = (x_s^2, x_o^2)$ is revealed worse than x^1 if it is within the budget set B, such that $x_s^2 + x_o^2 p \leq \acute{M}$. Substituting $\acute{M} = x_s^1 + x_o^1 p$ gives the condition that x^1 is revealed preferred to x^2 if $x_s^2 + x_o^2 p \leq x_s^1 + x_o^1 p$, unless the two points are equal ($x^1 = x^2$).

(*Appendices continue*)

To check WARP violations we follow Dobell (1965). Recapitulating, x^1 is directly revealed preferred to good vector x^2 if $p^1x^1 \geq p^1x^2$ and $x^1 \neq x^2$. In our setup, each good vector consists of the components x_s and x_o, such that $x^i = [x_s^i, x_o^i]$. Similarly, the price vectors are $p^i = [1, p_o^i]$ where $p_o^i = \frac{m_s^i}{m_o^i}$. We begin by constructing the matrix A such that $[a_{ij}] = [p^i \cdot q^j]$, where $(\cdot)$ represents the inner product. Thus, the entry in the first row, first column is:

$$a_{11} = p^1 \cdot q^1 = x_s^1 + p^1 x_o^1 \tag{3}$$

The entire matrix A can be constructed in the following way. Consider the vector of hold prices $p_s = 1$ and vector of good kept for the self x_s. Define $A_s = p_s \otimes x_s$, where $\otimes$ is the outer product of vectors. Then we have:

$$A_s = p_s \otimes x_s$$
$$= \begin{pmatrix} x_s^1 & x_s^2 & x_s^3 & x_s^4 & x_s^5 & x_s^6 & x_s^7 & x_s^8 \\ x_s^1 & x_s^2 & x_s^3 & x_s^4 & x_s^5 & x_s^6 & x_s^7 & x_s^8 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_s^1 & x_s^2 & x_s^3 & x_s^4 & x_s^5 & x_s^6 & x_s^7 & x_s^8 \end{pmatrix}$$

Similarly, define $A_o = p_o \otimes x_o$:

$$A_o = p_o \otimes x_o$$
$$= \begin{pmatrix} p_o^1x_o^1 & p_o^1x_o^2 & p_o^1x_o^3 & p_o^1x_o^4 & p_o^1x_o^5 & p_o^1x_o^6 & p_o^1x_o^7 & p_o^1x_o^8 \\ p_o^2x_o^1 & p_o^2x_o^2 & p_o^2x_o^3 & p_o^2x_o^4 & p_o^2x_o^5 & p_o^2x_o^6 & p_o^2x_o^7 & p_o^2x_o^8 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_o^8x_o^1 & p_o^8x_o^2 & p_o^8x_o^3 & p_o^8x_o^4 & p_o^8x_o^5 & p_o^8x_o^6 & p_o^8x_o^7 & p_o^8x_o^8 \end{pmatrix}$$

Then, define $A = A_s + A_o$:

$$A = \begin{pmatrix} x_s^1 + p_o^1x_o^1 & x_s^2 + p_o^1x_o^2 & \ldots & x_s^8 + p_o^1x_o^8 \\ \vdots & \vdots & \vdots & \vdots \\ x_s^1 + p_o^8x_o^1 & x_s^2 + p_o^8x_o^2 & \ldots & x_s^8 + p_o^8x_o^8 \end{pmatrix}$$

Thus, A has the desired property that $[a_{ij}] = [p^i \cdot q^j]$. We can then construct the matrix B that has the property $[b_{ij}] = [p^i \cdot q^j - p^i \cdot q^i]$. This can be done by subtracting a vector d with elements $d = [p^1 \cdot x^1, p^2 \cdot x^2, \ldots, p^8 \cdot x^8]$ from A, that is $B = A - d$. The vector d is just the diagonal elements of A as an 8×1 vector. Finally, we create a matrix C whose elements $c_{ij} = 1$ if $b_{ij} \leq 0$, and 0 otherwise. The matrix C represents the revealed preference relationship between each of the eight choices, where every entry that is equal to 1 implies $p^iq^j \leq p^iq^i$ meaning the row choice is directly revealed preferred to the column choice. The matrix C can also be described as an adjacency matrix. We then set to zero any cells where $x^i = x^j$, that is $[x_s^{ij}, x_o^{ij}] = [x_s^{ji}, x_o^{ji}]$.

(*Appendices continue*)

Appendix C

Checking SARP

Consider the direct revealed preference relations $A > B$, $B > C$, and $C > A$. Clearly this violates SARP. How do we construct an algorithm to figure this out? Consider the adjacency matrix R^0 that represents only direct revealed preference:

$$R^0 = \begin{pmatrix} & A & B & C \\ A & 0 & 1 & 0 \\ B & 0 & 0 & 1 \\ C & 1 & 0 & 0 \end{pmatrix}$$

For the adjacency matrix, there is a 1 in cell i, j if the option in row i is preferred to the option in column j. For example, cell $i = 1$, $j = 2$ is 1 because A is directly revealed preferred to B. We then take R^0 and make a sequence of adjustments that add indirect revealed preference relationships to this matrix. As we iterate through the algorithm, all of the indirect revealed preference relationships are also represented in the matrix. This is done, one at a time, by asking whether each option is indirectly revealed preferred to every other option, through each other option as an intermediary. This final matrix R^n, where $n = i \times j \times k$, is called the transitive closure of R.

The algorithm works using the following recursive update (Warshall, 1962):

$$R^n[i,j] = \begin{cases} 1, & \text{if } R^{n-1}[i,k] = R^{n-1}[k,j] = 1 \\ R^{n-1}[i,j], & \textit{otherwise} \end{cases} \quad (4)$$

Applying this to our matrix R^0, we start with $k = 1$, $i = 1$, and make the following adjustments by looping through $j = \{1, 2, 3\}$:

- $R^1[1, 1] = 1$ if $R^0[1, 1] = R^0[1, 1] = 1$
- $R^2[1, 2] = 1$ if $R^1[1, 1] = R^1[1, 2] = 1$
- $R^3[1, 3] = 1$ if $R^2[1, 1] = R^2[1, 3] = 1$

The first says that A is preferred to A if A is preferred to A and A is preferred to A. The second says that A is preferred to B if A is preferred to A and A is preferred to B. The third says that A is preferred to C if A is preferred to A and A is preferred to C. These updates make no change to the matrix, giving:

$$R^3 = \begin{pmatrix} & A & B & C \\ A & 0 & 1 & 0 \\ B & 0 & 0 & 1 \\ C & 1 & 0 & 0 \end{pmatrix}$$

Next, for $i = 2$, $k = 1$:

- $R^4[2, 1] = 1$ if $R^3[2, 1] = R^3[1, 1] = 1$
- $R^5[2, 2] = 1$ if $R^4[2, 1] = R^4[1, 2] = 1$
- $R^6[2, 3] = 1$ if $R^5[2, 1] = R^5[1, 3] = 1$

The first says that B is preferred to A if B is preferred to A and A is preferred to A. The second says that B is preferred to B if B is preferred to A and A is preferred to B. The third says that B is preferred to C if B is preferred to A and A is preferred to C. Again, none of these change the matrix.

$$R^6 = \begin{pmatrix} & A & B & C \\ A & 0 & 1 & 0 \\ B & 0 & 0 & 1 \\ C & 1 & 0 & 0 \end{pmatrix}$$

Next, for $i = 3$, $k = 1$:

- $R^7[3, 1] = 1$ if $R^6[3, 1] = R^6[1, 1] = 1$
- $R^8[3, 2] = 1$ if $R^7[3, 1] = R^7[1, 2] = 1$
- $R^9[3, 3] = 1$ if $R^8[3, 1] = R^8[1, 3] = 1$

(*Appendices continue*)

The first says C is preferred to A if C is preferred to A and A is preferred to A. The second says C is preferred to B if C is preferred to A and A is preferred to B. This case is true so we change $R^8[3, 2] = 1$. The third says C is preferred to C if C is preferred to A and A is preferred to C.

$$R^9 = \begin{array}{c} \\ A \\ B \\ C \end{array}\begin{array}{ccc} A & B & C \\ \end{array}$$

$$R^9 = \left(\begin{array}{c|ccc} & A & B & C \\ A & 0 & 1 & 0 \\ B & 0 & 0 & 1 \\ C & 1 & 1 & 0 \end{array}\right)$$

Thus, with $k = 1$, we have found all preference relationships between A, B, and C through A. With $k = 2$ we repeat this but try to find all preference relationships between A, B, and C through B. For $i = 1$, $k = 2$:

- $R^{10}[1, 1] = 1$ if $R^9[1, 2] = R^9[2, 1] = 1$
- $R^{11}[1, 2] = 1$ if $R^{10}[1, 2] = R^{10}[2, 2] = 1$
- $R^{12}[1, 3] = 1$ if $R^{11}[1, 2] = R^{11}[2, 3] = 1$

The first says A is preferred to A if A is preferred to B and B to A. The second says A is preferred to B if A is preferred to B and B is preferred to B. The third says A is preferred to C if A is preferred to B and B is preferred to C. The third one is true, so $R^{12}[1, 3]$ will be changed to 1.

$$R^{12} = \left(\begin{array}{cccc} & A & B & C \\ A & 0 & 1 & 1 \\ B & 0 & 0 & 1 \\ C & 1 & 1 & 0 \end{array}\right)$$

Next, for $i = 2$, $k = 2$:

- $R^{13}[2, 1] = 1$ if $R^{12}[2, 2] = R^{12}[2, 1] = 1$
- $R^{14}[2, 2] = 1$ if $R^{13}[2, 2] = R^{13}[2, 2] = 1$
- $R^{15}[2, 3] = 1$ if $R^{14}[2, 2] = R^{14}[2, 3] = 1$

The first asks whether B is preferred to B and B is preferred to A is connected to B and B is connected to A. The second is whether there is a path from B to B and B to B. The third is whether there is a path from B to B and B to C. None of these are relevant because B is not preferred to B.

$$R^{15} = \left(\begin{array}{cccc} & A & B & C \\ A & 0 & 1 & 1 \\ B & 0 & 0 & 1 \\ C & 1 & 1 & 0 \end{array}\right)$$

Then for $i = 3$, $k = 2$:

- $R^{16}[3, 1] = 1$ if $R^{15}[3, 2] = R^{15}[2, 1] = 1$
- $R^{17}[3, 2] = 1$ if $R^{16}[3, 2] = R^{16}[2, 2] = 1$
- $R^{18}[3, 3] = 1$ if $R^{17}[3, 2] = R^{17}[2, 3] = 1$

The first says C is preferred to A if C is preferred to B and B is preferred to A. The second says C is preferred to B if C is preferred to B and B is preferred to B. The third says that C is preferred to C if C is preferred to B and B is preferred to C, both of which are true so $R^{18}[3, 3] = 1$.

$$R^{18} = \left(\begin{array}{cccc} & A & B & C \\ A & 0 & 1 & 1 \\ B & 0 & 0 & 1 \\ C & 1 & 1 & 1 \end{array}\right)$$

Then for $i = 1$, $k = 3$:

- $R^{19}[1, 1] = 1$ if $R^{18}[1, 3] = R^{18}[3, 1] = 1$
- $R^{20}[1, 2] = 1$ if $R^{19}[1, 3] = R^{19}[3, 2] = 1$
- $R^{21}[1, 3] = 1$ if $R^{20}[1, 3] = R^{20}[3, 3] = 1$

The first says A is preferred to A if A is preferred to C and C is preferred to A, which is true, so $R^{19}[1, 1] = 1$. The second says A is preferred to B if A is preferred to C and C is preferred to B, but A is already preferred to B, so no change is made. The third says A is preferred to C if A is preferred to C and C is preferred to C, which again is already true.

(*Appendices continue*)

$$R^{21} = \begin{array}{c|ccc} & A & B & C \\ \hline A & 1 & 1 & 1 \\ B & 0 & 0 & 1 \\ C & 1 & 1 & 1 \end{array}$$

Finally, for $i = 2$, $k = 3$:

- $R^{22}[2, 1] = 1$ if $R^{21}[2, 3] = R^{21}[3, 1] = 1$
- $R^{23}[2, 2] = 1$ if $R^{22}[2, 3] = R^{22}[3, 2] = 1$
- $R^{24}[2, 3] = 1$ if $R^{23}[2, 3] = R^{23}[3, 3] = 1$

The first says B is preferred to A if B is preferred to C and C is preferred to A, which is true so $R^{22}[2, 1] = 1$. The second says B is preferred to B if B is preferred to C and C is preferred to B, which is true, so $R^{23}[2, 2] = 1$. Because the matrix is now all ones, no further updates are possible, and we finish the algorithm.

$$R^{24} = \begin{array}{c|ccc} & A & B & C \\ \hline A & 1 & 1 & 1 \\ B & 1 & 1 & 1 \\ C & 1 & 1 & 1 \end{array}$$

The resulting matrix R^{24} is the transitive closure of R^0, which shows all revealed preference relationships (direct and indirect). We can use this to calculate the SARP violations by looking at any case where $R^{24}[i, j] = R^{24}[j, i] = 1$ for $i \neq j$. This reveals that there are three SARP violations, of 1) $A > B$ and $B > A$, 2) $A > C$ and $C > A$, and 3) $B > C$ and $C > A$.

Appendix D

Checking GARP

GARP states that if x is revealed preferred to y (directly or indirectly) then $py \leq px$. We can check GARP by finding the transitive closure of x, then if xRy and $py > px$, GARP is violated. GARP requires only one additional step above SARP. After computing the transitive closure of each choice to produce the matrix R^n, we evaluate whether it is true that $R^n[i, j] = R^n[j, i] = 1$, which would indicate that alternative i is (directly or indirectly) preferred to j and alternative j is preferred to i. We then construct a matrix G^n by modifying R^n to $G^n[i, j] = 0$ if $G^n[i, j] = G^n[j, i] = 1$ and $p^i p^j = p^i p^i$.

Appendix E

Experimental Procedures

Participants entered the lab, took a seat, and were instructed to carefully read the informed consent document. The consent ensured that no participant had any prior medical issues, such as frostbite, that would interfere with their participation in the experiment. No potential participants refused to participate. Participants were then randomly assigned an identification number unknowable to others using shuffled opaque sealed envelopes. Numbers were chosen such that if there were an odd number of participants, the unpaired person would be assigned to be a receiver, so that all dictators' decisions would be real.

(*Appendices continue*)

After random assignment, initial instructions were read aloud to the participants. Participants were told that they would be paid $10 in cash at the end of the experiment. Participants were informed that they would be split into two rooms (based on whether they had an even or odd identification number), and that each odd-numbered participant (dictator) would make a decision that would affect an even-numbered participant (receiver), but that even-numbered participants would not make any choices that affected the odd-numbered participants. They were asked not to communicate with others and to raise their hand if they had any questions during the experiment. All participants were assured that their choices throughout the experiment would be anonymous.

There were three experimenters, a female who always stayed with the receivers, a male who stayed with the dictators, and a third male in an additional room to deliver payoffs. The two experimenters with the receivers and dictators were both blind to the experimental hypotheses. Before making their choices, participants came to the "Ice Water Experience" section of the questionnaire, where they were asked to raise their hand to notify the experimenter. The experimenter then walked to the participant with an insulated container of ice-water maintained at 3–5 degrees Celsius, and had the participant submerge her hand in the water for 5 s, timed with a digital stopwatch that was observable by the participant.

Dictators then made eight main allocation decisions for their paired receiver. All participants then completed an auction to determine how much they valued avoiding the ice-water experience. When all participants were finished, a random number generator picked one of the eight decisions to be implemented. The outcome of this decision was then recorded on a slip of paper for the associated dictator and receiver and placed in individual envelopes, though the experimenter could not attribute choices to any particular individual in the room. These were given to the third experimenter in the payment room, in such a way that he could only know the amount of immersion time to give each participant, but could not identify any participant's role in the experiment.

Once this procedure was finished, the receivers were called into the payment room, one at a time, and received their payoffs. Receivers were asked to submerge their hand in the ice-water container for the duration of time specified by the associated envelope. The participant's auction decision was then checked to see if she had won the auction, and if so, her hand was submerged in ice water for 60 s and she was paid accordingly. Finally, they were paid their $10 show-up fee. This process was repeated for dictators, who put their hand in the ice-water for the duration of the time specified by their randomly selected choice and then received the outcome of the auction. Although participants were told in the informed consent agreement that they could stop at any point and that doing so would not result in any penalty or loss of benefits, no one failed to complete the task or dropped out.[11]

[11] There is the potential that participants might have been confused about whether discontinuing participation would forfeit their show-up fee (this was not the case), though we have no evidence that this occurred.

(*Appendices continue*)

Appendix F

Price Elasticities

As a final measure of overall choice behavior, we calculate the arc price elasticities of demand for the good, time for self not in water. The arc price elasticity measures the percent change in the amount of time for self not in water relative to a percent change in the price of time for self not in water. In this case, the dictator pays this price by putting the other person's hand in ice water for some amount of time, so the price of giving is captured in p.[12] For example, consider two budget sets with $\acute{M} = 60$ and $p = \{1, 2\}$. Here, $p = 1$ indicates that for each second of time for other not in water, the dictator must lose 1 s of time for self not in water. Similarly, for $p = 2$, allocating 1 s of time for other not in water incurs a 2 s loss of time for self not in water. Thus, if the dictator changes the amount of time for self not in water from 60 to 50 when the price changes from 1 to 2, the proportionate change in time for self not in water is -0.18 ($= \frac{50 - 60}{(50 + 60)/2}$), indicating that the amount of time for self not in water dropped by 18%. The proportionate change in price is 0.66 ($= \frac{2 - 1}{(2 + 1)/2}$). Thus, the arc price elasticity is -27% ($= -100 \times 0.18/0.66$), indicating that as the price of giving time for other not in water increases, the amount of time for self not in water decreases, which is the opposite of what would normally be expected, where less of a good is purchased when its price increases.

Arc price elasticities can be calculated using choices across budgets where prices vary but incomes do not. There are four such budget sets in our experiment (with $\acute{M} = 60$, and $p = \{1, 1.5, 2, 3\}$). If our dictators had perfectly selfish preferences, these elasticities would all be 0 (because, the numerator would always be zero for a selfish person, who would keep all of the time not in water regardless of the price of giving). If, instead, their preferences were pure Leontief, the respective elasticities would equal -0.56, -0.64, and -0.71, with an overall mean of -0.64.[13]

In our experiment, the observed mean elasticities for the three separate price increases and overall are given in Table F1. The average elasticity across all price changes is -0.40, indicating that as the price of giving time for other not in water increases, the average amount of time for self not in water decreases, which is the opposite of normal demand patterns. Of the 27 dictators, 19 (70%) of them had nonpositive elasticities for all three price changes, and out of the 81 (3×27) elasticities observed, only 9 (11%) were positive. These results differ from prior work in the domain of positive, monetary allocations where, for example, calculated equivalent arc elasticities were found to have the opposite sign (Andreoni & Vesterlund, 2001). The negative demand elasticities we observe in our participants are consistent with equality-seeking behavior, rather than efficiency-seeking or social-welfare maximization.

[12] Technically p is the price of giving time not in water to the other person, paid by the dictator putting her hand in water. However, for arc price elasticities the same result obtains if p or its inverse $\frac{1}{p} = \frac{m_o}{m_s}$ (the price of keeping time not in water for self) is used.

[13] Someone with pure Leontief preferences would allocate tokens in these three situations such that $y_s = y_o$. Because $y_s = Km_s$, and $y_o = (M - K)m_o$, a little algebra shows that $K = \frac{M}{p+1}$. As a result, the amount of bad kept is $y_s = m_s\frac{M}{p+1}$ and the mount of good kept is $x_s = 60 - m_s\frac{M}{p+1}$. Using this, a dictator with pure Leontief preferences will split the tokens such that both dictator and receiver get 30 s ($p = 1$), 36 s ($p = 1.5$), 40 s ($p = 2$), and 45 s ($p = 3$) of the bad. The equivalent amount of good kept by the dictator is 30, 24, 20, and 15 s, respectively. Thus, going from $p = 1$ to $p = 1.5$ incurs a change in the quantity of good kept (x_s) from 30 to 24, giving a numerator of -0.22 $\left(= \frac{24-30}{(30+24)/2}\right)$ and a denominator of 0.40 $\left(= \frac{1.5 - 1}{(1.5 + 1)/2}\right)$, resulting in an arc price elasticity of -56% $\left(= 100 \times \frac{-0.22}{0.40}\right)$. The other arc price elasticities for a dictator with Leontief preferences would be -0.64 ($p = 1.5$ to $p = 2$), and -0.71 ($p = 2$ to $p = 3$), with an overall mean of -0.64.

Table F1
Average Arc Elasticities Given the Observed Choices of the Dictators (With $\acute{M} = 60$)

Price of giving	$1 \rightarrow 1.5$	$1.5 \rightarrow 2$	$2 \rightarrow 3$	Overall
Average arc elasticity	−.53	−.45	−.31	−.43
Choice pairs with arc elasticity > 0	0%	15%	19%	11%

Appendix G

Critical Cost Efficiency Index (CCEI)

Recall that if x_1 is chosen given budget set **B1** : $\acute{M} = x_s^1 + x_o^1 p^1$, then x^1 is revealed preferred to x^2 if $x_s^2 + x_o^2 p^1 \leq x_s^1 + x_o^1 p^1$, unless the two points are equal ($x^1 = x^2$). If it is also the case that given $\mathbf{B_2}$: $\acute{M} = x_s^2 + x_o^2 p^2$ and a choice of x^2 it happens to be true that $x_s^2 + x_o^2 p^2 \geq x_s^1 + x_o^1 p^2$, then x^2 is revealed preferred to x^1, violating WARP. If the inequality is strict in at least one of the two choices, then the preferences violate GARP.

To avoid this violation, we can multiply each budget set by a number e such that $x_s^2 + x_o^2 p^1 \leq e \times (x_s^1 + x_o^1 p^1)$ and $e \times (x_s^2 + x_o^2 p^2) \geq x_s^1 + x_o^1 p^2$ removes any WARP/GARP violations. e is called the critical cost efficiency index for a participant, and measures how much of the participant's endowments $\acute{M}^i$ must be thrown away to make the participant's choices consistent.

Table G1 shows the distribution of CCEIs across the 27 dictators. In terms of CCEI, the most direct methodological comparison with our experimental design is that of Andreoni and Miller (2002). The actual budget sets, and observed heterogeneity of preferences (discussed below), differ between the two experiments. To account for this, we calibrated the two experiments by assuming that participants with the two observed distributions of prototypical preferences made choices with some error, on the two different budget sets used. We found that, conditional on the budget set, there was little difference in the resulting mean CCEI for the two distributions of participants, with the budget set used in our experiment leading to a mean CCEI of 0.95 and that used in Andreoni and Miller (2002) of 0.93, across either distribution of preference types. Thus, the differences in rationalizability we observe across the two experiments are not likely driven by differences in the underlying budget sets or observed preference distributions.

Table G1
Distribution of the CCEI Across the 27 Dictators

CCEI	Number of subjects	Cumulative subjects	Cumulative %
1.000	14	14	52%
.999	5	19	70%
.979	1	20	74%
.917	1	21	78%
.833	1	22	81%
.749	3	25	93%
.494	1	26	96%
.333	1	27	100%

Note. CCEI = Critical Cost Efficiency Index.

Fisman et al. (2007), also in the realm of monetary gains, found 46% (35/76) of their two-person-condition participants had such violations, however, their study involved fifty, randomly generated budget sets. If participants made random choices on the three respective budget sets, we would predict 26% of participants would have severe CCEI in our study, 37% in Andreoni and Miller (2002), and 100% in Fisman et al. (2007).

Finally, by using an individual's CCEI measure, we can determine if the degree of GARP violation is related to observed behavior. We find no significant relationship between the mean amount of time not in water kept and CCEI ($r = -.08$). The mean WTA of the seven dictators with CCEIs below 0.95 is \$4.16, while the mean for the 19 (eliminating the outlier) with higher CCEIs is \$2.89, though the difference here is not statistically significant.

Received March 8, 2016
Revision received December 13, 2016
Accepted December 29, 2016 ■