What you don't know won't hurt me: Costly (but quiet) exit in dictator games

Jason Dana [a,*], Daylian M. Cain [b], Robyn M. Dawes [b]

[a] Department of Psychology, University of Illinois, Urbana-Champaign, 603 E. Daniel St. MC 716, Champaign, IL 61820, USA
[b] Carnegie Mellon University

Abstract

We used simple economic games to examine pro-social behavior and the lengths that people will take to avoid engaging in it. Over two studies, we found that about one-third of participants were willing to "exit" a $10 dictator game and take $9 instead. The exit option left the receiver nothing, but also ensured that the receiver never knew that a dictator game was to be played. Because most social utility models are defined over monetary outcomes, they cannot explain choosing the ($9, $0) exit outcome over the dominating $10 dictator game, since the game includes outcomes of ($10, $0) and ($9, $1). We also studied exiting using a "private" dictator game. In the private game, the receiver never knew about the game or from where any money was received. Gifts in this game were added innocuously to a payment for a separate task. Almost no dictators exited from the private game, indicating that receivers' beliefs are the key factor in the decision to exit. When, as in the private game, the receivers' beliefs and expectations cannot be manipulated by exit, exit is seldom taken. We conclude that giving often reflects a desire not to violate others' expectations rather than a concern for others' welfare per se. We discuss the implications of our results for understanding ethical decisions and for testing and modeling social preferences. An adequate specification of social preferences should include "psychological" payoffs that directly incorporate beliefs about actions into the utility function.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Fairness; Altruism; Dictator game; Psychological games

Introduction and outline

An enduring question in the social sciences is whether generosity exists without underlying selfish motives. For example, if someone gives a substantial sum to charity should we infer that the donor truly cares about the charity's cause? Perhaps not—the donor might be giving money to get a tax break, to impress onlookers, to inspire quid pro quo from those who receive the donation, or for a myriad of other reasons. One way to shed light on motives for giving is to use economic games in a laboratory to control these sorts of confounds. For example, in the "dictator game," one participant—the dictator—is given an endowment of money that she may divide however she likes with an anonymous other participant—the receiver. The receiver, who typically knows the dictator's instructions, must accept whatever division the dictator makes (even if given nothing) and cannot punish or retaliate. The dictator game is typically run in one-shot fashion; participants will not get to exchange roles or play again. Thus, the structure of the game rules out strategic incentives for giving.

A rational dictator who cares only about maximizing selfish payoffs should give nothing to the receiver.

* Corresponding author.
E-mail address: dana@cyrus.psych.uiuc.edu (J. Dana).

However, across several studies using the dictator game, dictators give a mean amount of between 20 and 30% of the endowment, and often a majority of dictators give positive amounts (for a review, see Camerer, 2003). Because strategy is not a concern, one might conclude that dictators who give do so because they care about the receiver's welfare. In fact, as we will discuss, many current "social preference" models essentially postulate just that. This conclusion, however, does not necessarily follow from the fact that dictators give. We will argue that at least some giving occurs because the dictator does not want to appear selfish, even to an anonymous receiver who cannot punish the dictator. More specifically, we posit that many dictators are motivated to give what they think receivers expect them to give. For this reason, many dictators will give to receivers in situations like the dictator game, where it is known that they have been given an endowment and that their actions are transparent to the receiver. However, if these same dictators can prevent the receiver from ever knowing about the game, even at a small cost, they may do so because the receiver will then expect nothing and the dictator will not have to give. Such behavior, which is akin to crossing the street to avoid a beggar, is inconsistent with the interpretation that the dictator gives because she cares about the receiver's welfare.

We present two original studies that demonstrate that dictators may give without any true concern for the receiver's welfare. In study 1, participants in the role of dictator were asked to allocate $10 in the standard dictator game framework. After making their choices, but before receivers were told about the game, dictators were given the option of 'exiting' for $9 instead of playing the game. If the exit option was chosen, receivers were given nothing but were also not told about the game. A substantial minority of dictators were willing to exit, in spite of the fact that the $10 game dominates the exit option in terms of monetary payout.

In a second study, a replication of study 1 again showed a substantial proportion of dictators exiting, and served as a baseline for comparison of exiting from a new "private" dictator game. In the private game, receivers were not told why they received any money given to them, so that they had no expectations of a gift no matter what the dictator did. Only one dictator exited from the private game, suggesting that exiting truly reflects a desire to avoid the receiver's expectations, rather than alternative explanations that we later discuss, including desires to limit one's own action set, escape responsibility, or simply comply with implicit experimenter demands.

The remainder of the paper is organized as follows. In the next section, we give a general discussion of current social preference models, focusing on how they explain giving in games and aspects of giving that they cannot explain. Then, we present the procedures and results of our two studies. Finally, we discuss the implications of our findings for the study of fairness and generosity and suggest modeling approaches that can better capture the sort of concerns we have described. Specifically, exit behavior is explained best by using "psychological payoffs" (Geanakoplos, Pearce, & Stacchetti, 1989) that directly incorporate decision makers' beliefs into their utility functions. We give an example in which the dictator's utility is given by her own payoff, discounted by a comparison of *her gift to the receiver* with *what she thinks the receiver expects her to give*.

Background

To reconcile giving with the traditional economic assumption of self-interest, several authors have proposed models of social preference. A class of social preference models that can be called distributional models is pertinent to behavior in the dictator game. These models describe preferences in terms of the distribution of payoffs to all parties. For example, inequity aversion or difference aversion models (e.g., Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999; Loewenstein, Thompson, & Bazerman, 1989) hold that as the discrepancy between outcomes for the various stakeholders increases, the decision maker experiences more disutility. While this disutility is often taken to be stronger for inequality favoring other parties over oneself, self-advantageous inequalities are taken to be dispreferred as well (Bethwaite & Tompkinson, 1996; Fehr & Schmidt, 1999; Kahneman, Knetsch, & Thaler, 1986; Loewenstein et al., 1989; Messick & Sentis, 1985; Messick, 1995). Thus, in dispute resolution and when allocating money in general, peoples' selfish interests are countervailed by their desires to arrive at equal outcomes. As such, the social comparison component is generally taken to reflect a "taste for fairness."

Various other formulations of distributional social preference include concerns for efficiency (defined as the sum of all payoffs) and maximizing the smallest payoff to any party ("maximin preferences," as in Charness & Rabin, 2002; Engelmann & Strobel, 2004); or they hold that charity is motivated by the selfish good feelings it engenders (e.g., Andreoni, 1990), making giving a self-interested choice. In fact, over repeated dictator games, it was found that participants' choices obey rational choice axioms when "others' outcomes" are treated as any other good that people consume (Andreoni & Miller, 2002).

While distributional models are often intended to implicitly capture psychological aspects of fairness and giving, they all explain fairness solely in terms of preferences over payoffs. Thus, dictators are assumed to give because they like the more equitable outcome. Assuming a taste for fair outcomes seems sensible because generosity

occurs in economic experiments that ostensibly control for ulterior motives to give. Yet, there may be other important motives for giving that we cannot adequately capture with payoffs alone, for example the desire not to be seen as selfish. It is this motivation that is the concern of the experiments in the next section.

Gaertner (1973) gave a classic demonstration of how a concern for appearances may lead to altruistic behavior, even though the target of generosity is anonymous and cannot sanction. An African American confederate (using a voice and accent which suggested this ethnicity) called white liberals and conservatives claiming that he dialed their number by mistake with his last dime and asked them to call a tow truck for him. Gaertner found that liberals were more likely to help if asked, but that they were also more likely to hang up before the request could be made. Being willing to help if asked but hanging up the phone early could be construed as refusing information about whether the caller needed help and as such, cannot be explained by a coherent preference over outcomes alone. If liberals really cared about the caller's welfare, they would want to know if the caller needed help. Gaertner concluded that the liberal participants helped more because they were more concerned with not appearing racist by refusing the caller's request once it was made.

Similarly, our experiments will show how a concern for not appearing selfish can sometimes motivate generosity, but a kind of generosity that participants would rather have avoided. As such, our results cannot be explained by any model defined over payoff distributions. Of course, others have pointed out problems with distributional models. For example, some have argued that seeming difference aversion is "intention-based" reciprocity, stemming from a desire to be kind to those who have been kind and spiteful to those who have not (e.g., Dufwenberg & Kirchsteiger, 2004; Rabin, 1993). For instance, experiments have shown that people are not as averse to unequal outcomes that are generated by a computer (e.g., Blount, 1995) or when the allocating party cannot choose an equal outcome (Falk, Fehr, & Fischbacher, 2003). In these cases, unequal outcomes are more palatable, apparently because they do not reflect that another person has been unkind.

Intentions-based models cannot explain a dictator's generosity, however, because the receiver is passive. That is, the receiver has done nothing from which the dictator can infer his intentions. Further, our experiments point out a more fundamental flaw in the analysis of prosocial behavior in experiments. In terms of monetary outcomes alone, our dictators make self-contradictory choices that violate the mandates of decision theory. Thus, we point to the need to identify a missing factor in the analysis, one that leads people to give when they do not "have to." Specifically, we argue the need to

consider the dictator's desire to behave as she thinks the receiver expects her to behave. Some support for this idea comes from Dufwenberg and Gneezy (2000), who found that participants' reports of what they think others expect to receive are positively correlated with the amounts they give. We seek to extend such findings by allowing the dictator to manipulate the receiver's expectations, thus establishing that these expectations are a causal influence on giving.

We are not claiming that there are no genuine concerns for others' welfare. Indeed, many of our dictators give even after being given the opportunity to escape scrutiny by exiting. However, interpretations of prior experimental results may be overstating the level of concern for others' welfare that people have. Sometimes people give (even in non-strategic contexts) without ultimately caring about the recipient's welfare. Instead, they have an intrinsic desire to merely meet others' expectations.

To cast this motivation in an economic game example, a dictator who thinks that the receiver expects her to split a $10 endowment may feel compelled to give $5 due to her desire to appear fair. But if appearances are the reason the dictator shares, she may wish to keep her endowment private so that the receiver does not expect anything. Then, she does not have to give or feel guilty for not giving. If she is sensitive enough to the receiver's expectations, she may even be willing to accept $9 if she can keep the receiver from knowing about the game. This behavior is reasonable given a concern for appearances, yet it cannot be explained by any preference defined only over outcomes because the outcome of $9 for the dictator and $0 for the receiver (henceforth denoted ($9, $0)), is dominated by the dictator game, which includes outcomes of ($9, $1) and ($10, $0).

One may question why a dictator would care about what the receiver thinks if the receiver is anonymous and cannot punish. Perhaps one reason is that across a broad spectrum of situations, a concern for appearing fair is evolutionarily advantageous in that it fosters beneficial cooperation and helps avoid sanctions (Bowles & Gintis, 2004; Fehr & Gachter, 2002; Gintis, Bowles, Boyd, & Fehr, 2003). Because it is so often useful, a concern for appearances can quickly proliferate in a population. Yet, this concern is probably more automatic than reflective, and we cannot simply turn it off in those rare situations where it serves no strategic purpose, as when we act with true anonymity. Elster (1989) gives us a striking example: most people would not pick their noses if they could be seen by strangers on a passing train, even if those passengers are effectively anonymous because they will never be seen again.

If our hypothesis is accurate—that many apparently fair dictators are indeed only interested in appearing fair—then we should see some dictators accept $9 in

private instead of a $10 dictator game. The reasoning is simple. A $0 outcome to a receiver who knows nothing of a dictator game is decidedly neutral; we would not expect others to have feelings about the fact that money has not fallen from the sky. Thus, exiting allows these dictators to be selfish without feeling that they have disappointed the receivers. Put differently, such dictators would behave as if "What you don't know won't hurt you, and ergo, won't hurt *me*." With this idea as motivation, we conducted a study that allowed dictators to exit a $10 dictator game, i.e., keep it secret from the receiver, for $9.

Study 1

To test our idea, we used a $10 Dictator game with an exit option. This exit option allowed the dictator to accept $9 to end the experiment without the receiver getting any money or learning about the foregone Dictator game.

Methods

Participants and design

Participants were 80 students (40 dictators) at Carnegie Mellon University recruited through campus advertising to participate in paid experiments. Data were collected in 4 sessions with at least 18 participants present in each session. The roles of dictator and receiver were assigned randomly and kept anonymous. Study 1 was administered silently following an unrelated task in which all participants filled out a series of surveys and were paid $10. Because study 1 was unannounced and participants were already in the room completing another task, the dictator could exit without the would-be receiver suspecting that a dictator game was to have taken place.

Procedure

Participants were seated apart in a large room. Dictators were given written instructions about the game for both themselves and the anonymous receiver. The words dictator and receiver were not used in the instructions to describe the roles; both parties were referred to simply as participants. Dictators were asked to allocate $10 of real money between themselves and the other participant in $1 increments (10–0, 9–1, 8–2, etc.). The dictator's instructions explained that the experimenter would discretely transfer both sets of instructions to the other participant once the allocation decision had been made, even if the dictator gave nothing. The receiver's instructions explained the game and stated that the dictator's instructions were included. In this way, the participants would have common knowledge if the game

was played: The receiver would know what the dictator knew when making the choice, and the dictator would know that the receiver would know what the dictator knew, etc. No mention of the exit option was made at this point.

After dictators made their putative allocation choices, but before any instructions were transferred, they were given a sheet of paper asking them to make a second choice. Dictators were offered $9 instead of their allocation of the $10. If dictators chose the $9 exit option, they would keep both their own and the receiver's instruction sheets. The receiver would be given nothing, and would not be told about the dictator game. If dictators chose to stay with the dictator game, their original division of the $10 would be executed and the instruction sheets would be transferred to the other participant. Regardless of which option was chosen, the task ended and the dictator was dismissed. Choosing the exit option did not get the dictator out of the experiment any faster than staying with the initial allocation. Study 1 was always administered at the end of the experimental sessions, and all participants were paid privately in cash as they exited the room.

Any model that defines preferences over monetary outcomes predicts that dictators will not exit, because the outcome ($9, $0) is dominated by the dictator game. If the dictator is selfish, keeping all $10 must be better than accepting $9. If the dictator cares about the receiver's welfare, keeping $9 and giving $1 must be better than the exit, and other dictator allocations may, in turn, be better than giving away only $1. If we assume, however, that dictators care about how their actions appear to the anonymous receiver, exiting may be coherent.

For example, suppose that a dictator does not truly care about the receiver's outcome, but does care about appearing selfish to the receiver.[1] That concern may cause the dictator to share some of the endowment, perhaps more than $1. By exiting, the dictator no longer has to worry about the appearance of selfishness because the receiver will not know anything about the endowment, and the $9 is more than she would otherwise be comfortable keeping. Alternatively, the dictator may not be concerned enough with appearing fair to share the endowment. Perhaps the dictator believes that the receiver expects $5 from a $10 endowment and does not value the appearance of generosity enough to give $5. Such a dictator may give nothing. Yet, if she values appearances at all, the same dictator may exit, avoiding the appearance of selfishness for only $1.

---

[1] A truly generous dictator could also be concerned with appearing foolish if she thinks that others expect her not to give (e.g., if she thinks people believe the "smart thing to do" is to give nothing), a topic we develop in the general discussion section.

We thus predict that a substantial proportion of dictators will choose the ostensibly dominated exit option. Because a dictator that worries about the receiver's beliefs may exit after either sharing or keeping everything, as described above, we expect little relationship between the size of the initial gift to the receiver and the decision to exit.

Study 1 Results

Allocation decisions

A histogram of dictators' putative gift amounts in study 1 is given in Fig. 1. This distribution of gifts is similar to those from previous dictator experiments in which roles were assigned randomly and experimenter-blind procedures were not used (see review in Camerer, 2003). Specifically, no dictators made ''hyperfair'' offers of more than half of the endowment, 30% of dictators offered half of the endowment, 33% offered nothing, and the remaining dictators offered amounts in between with roughly equal frequency. The mean gift amount was $2.40.

Exit decisions

Eleven of the 40 dictators (28%) took the exit option, 2 of which had intended to keep all $10. While the percentage of dictator exits that constitutes interesting behavior might be arbitrary, 28% seems more than ''trembles'' or noise; genuine preferences for exiting probably exist. This finding alone is problematic for theories of fairness in games. Fig. 1 (in the dark portion of the bars) shows actual gifts given by dictators who did not exit. As can be seen, the distribution is nearly identical to the distribution of putative gifts (in the total portion of the bars), supporting our prediction that the exit decision would relate little to gift size. The mean gift amount of dictators not choosing to exit was slightly lower at $2.28, but the correlation between putative gift amounts and exiting was small and not significant ($r = .10$, $p = .54$).
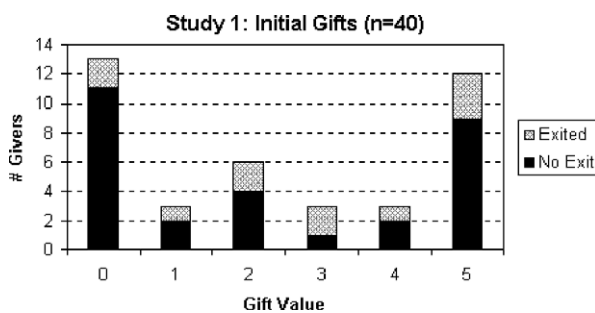

Fig. 1. Histogram of initial gifts in study 1.

Study 1 discussion

Several dictators exited, even though it meant accepting $1 less than they could have given themselves. Further, exiting cannot be said to have allowed the dictators to avoid thinking through the difficult decision of how to divide the $10, because they had to think through this decision before they knew about the exit option. Thus, exiting also meant abandoning their earlier decisions. Because the exit option is dominated in terms of monetary outcomes, neither selfishness nor social utility models predict that it should be chosen. These results support our contention that some people give because they are concerned with appearing fair to the recipient, but would rather keep the recipient blind so that they do not have to give.

However, there are alternative explanations for our findings that study 1 cannot rule out. We are aware of at least three:

(1) It could be that dictators view the exit as an escape from responsibility for the receiver. Although they might ultimately be responsible for the receiver getting nothing by exiting, participants may not frame the problem in this way. In this case, exiting could reduce intrapsychic conflict for dictators apart from any concern about what the receiver thinks.
(2) Exiting changes the dictator's action set; the dictator now has $9 that cannot be shared, so there are no choices to make. The dictator may value the situation in which sharing is not possible apart from any concern about what the receiver thinks.
(3) Exiting may be a type of experimenter demand effect. That is, the fact that such an option is offered may indicate that the experimenter thinks it should be used. Thus, the fact that some dictators exited may mean nothing about the properties of the exit, but rather that an option will always be chosen by some participants simply because it is offered.

To rule out these alternative explanations and more directly test the psychological mechanism we propose, we constructed a second study. In study 2, dictators were asked to divide the $10 with receivers who would not be told why they received any gift they were given. After making this decision, dictators were given the same $9 exit option. If dictators exit from this new game, then clearly exiting cannot be viewed as an escape from the receiver's expectations, since the receiver does not know an endowment is being divided, exit or no exit. However, if exiting is significantly reduced, we can rule out the above three explanations (exit to remove responsibility, to reduce action set, and to succumb to demand effects) because all of them apply generally to why dictators might exit from an allocation choice, whether that choice is known or unknown to the receiver. For

example, if dictators exit because they want to escape intrapsychic conflict or because they enjoy limiting their action sets, then they should still exit in the private condition because they face the same dilemma. If dictators exit simply because they are offered the choice, then they should exit at the same rate when offered the choice in the replication and the private conditions. Since we do not think these explanations drive exit, we predict that exit will be reduced in the private game.

Study 2

In study 2, we compare the exit rates of dictators from two different dictator games. In a *replication* condition, we use the same basic game as in the previous study: a standard $10 dictator game is played, and then a $9 exit option is offered (replicating the conditional common knowledge game). In a *private* condition, we use a dictator game in which the dictator's decision is private; the receiver does not know from where any money received has come. Because the receiver's knowledge is never a factor in the private condition and because we think that dictators exit because of concerns for what the receiver thinks, we predict that significantly fewer dictators will exit the private game than in the replication condition.

Methods

Participants and design

Ninety participants (45 dictators, 24 in the private condition) were recruited through campus advertising and from a research participation pool for course credit in the Tepper School of Business at Carnegie Mellon University. Students were eligible for the pool by taking business classes, but their major fields of study reflected diverse departments. Data were collected in 8 sessions (4 for each condition) ranging in size from 8 to 16 participants in the room. Participants were assigned to roles and conditions randomly. As in study 1, the dictator games were administered silently following unrelated tasks. In most sessions, participants received course extra credit and from $0 to $5 (mean of $3.50) for these unrelated tasks. In one session in each condition, they received a flat rate of $7. Because the preceding tasks were different from those used in study 1, and because study 1 occurred several months earlier, we replicated study 1 as a baseline against which to compare exits in the new private dictator game.

Procedure

Procedures were similar to those used in study 1, with the following exceptions: in the private condition, the dictator's instructions indicated that the other party would not be told about the game. The instructions explained that any money allocated to the other party would be accompanied by a note that read: "Thank you for your time today. An additional payment of $__ is attached to this note." Because this experiment was the last in a series of tasks and because payments were being given out for other tasks, we were able to distribute these payments discretely and receivers accepted them without question. After the allocation choice was made, the dictator was offered the same exit option as in study 1.

Study 2 Results

Allocation decisions

A histogram of the putative gifts amounts of dictators in both the known (replication) and private conditions is given in Figs. 2A and B, respectively. Dictators were slightly more generous in the replication condition of study 2 (mean gift = $2.67) than in study 1 (mean gift = $2.40), but this difference was not significant ($t = .47$, $p = .64$). Only 24% of dictators gave nothing, while 38% gave half of the endowment. In the private condition, the mean gift amount was $1.79, with 46% of dictators giving nothing and only 25% giving $5. These results support the contention that at least some giving occurs because the receiver expects it, but the difference in mean gifts between the private and replication conditions did not attain significance ($t = 1.36$, $p = .18$). We are not concerned that some giving occurs absent
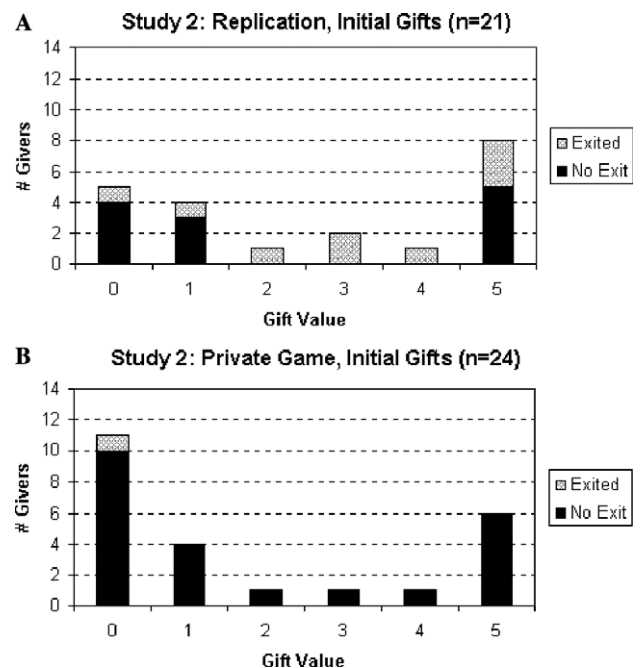

Fig. 2. Histograms of initial gifts in study 2, (A) replication and (B) private.

Table 1
Dictator behavior across conditions

	Initial giving > 0	Exit after giving > 0	Exit after giving 0	Total exit	No exit giving > 0	No exit giving 0
Study 1	27/40 (68%) $m = \$2.40$	9/27 (33%)	2/13 (15%)	11/40 (28%)	11/40 (28%)	18/40 (45%)
Study 2: replication	16/21 (76%) $m = \$2.67$	8/16 (50%)	1/5 (20%)	9/21 (43%)	4/21 (19%)	8/21 (38%)
Standard game all	43/61 (70%) $m = \$2.49$	16/43 (37%)	3/18 (16%)	20/61 (33%)	15/61 (25%)	26/61 (43%)
Study 2: private	13/24 (54%) $m = \$1.79$	0	1/11 (9%)	1/24 (4%)	10/24 (42%)	13/24 (54%)

expectations, however, because we do not doubt that some genuine preference for others' welfare exists. However, absent receiver expectations, we predict that choices should be more reflective of the dictator's true preference over outcomes, so that there should be no reason to abandon one's choice by exiting in the private condition.

Exit decisions

Table 1 lists the proportion of dictators who exited by condition (column 5). As can be seen, only 1 of the 24 dictators in the private condition (4%) chose to exit, while 9/24 (43%) exited in the replication condition. Consistent with our hypothesis that exiting reflects a concern about what the receiver thinks, the difference in exit proportions was significant (Pearson $\chi^2 = 9.70$, $p < .01$). Thus, dictators were much more likely to stick with their initial decisions when these decisions were made without the possibility of scrutiny by the anonymous receiver.

Exits were also more frequent in the study 2 replication (43%) than in study 1 (28%). This difference, however, was not significant (Pearson $\chi^2 = 1.47$, $p = .23$). Fig. 2A shows the histogram of actual gifts given by dictators who did not exit in the replication condition. Half of the dictators who had intended to give some positive amount (8 out of 16) were doing so "reluctantly" in that they exited from this choice. We find this behavior to be compelling because these dictators are essentially reneging on sincere intentions to give to others who are anonymous and cannot punish them. The mean gift amount for dictators not choosing to exit in the replication condition was $2.33, and again the correlation between putative gift amounts and exiting was not significant ($r = .18$, $p = .43$).

Study 2 discussion

In the replication condition of study 2, we found that a substantial proportion of dictators preferred to exit from the game. Many of our dictators made sincere commitments to give positive amounts without the threat of punishment or damaged reputation. Yet, half of those dictators chose to exit and leave the receiver with nothing, at a $1 cost to overall welfare. These results challenge the way that we currently model preferences for fairness in games. We will discuss how such preferences might be better captured in the next section.

In the private condition of study 2, exiting was nearly extinguished. The private condition differed from our other studies only in that the game from which the dictator could exit involved receiver blindness from the start. To our knowledge, this is the first experiment that uses this methodological variation on the dictator game. Yet, it seems a logical extension because the receiver is not a strategic player in the game and therefore should not require knowledge of strategies. The fact that dictators did not exit in this variation allows us to rule out several alternative explanations for exiting that do not involve a concern about what the receiver thinks. The results of study 2 strengthen our confidence that beliefs about the receiver's beliefs are the mechanism driving the choice to exit.

General discussion

Of all the dictators who had intended to give a positive amount, 40% (17/43) chose to renege their gifts by exiting, despite the fact that these gifts were made anonymously and under no threat of punishment. Including those who initially gave nothing, the overall exit rate from a standard dictator game across two studies was 20/61 (33%). The fact that 40% of givers were "insincere" in this manner shows that we cannot assume that a dictator who gives likes the fairer outcome. Further, because exiting was virtually non-existent when the endowment was initially private, it seems that the receiver's knowledge is the key factor in the dictator's decision to exit. We conclude that in addition to monetary payoff concerns, experimental participants are also motivated to take the action they think that others expect them to take. This conclusion has important implications for how we use games to study fairness and how we model fairness in games.

For example, if giving often reflects a concern for doing what one thinks others expect, it is important to choose a game in which there is a large consensus about

what actions are and are not expected. We used a dictator game here because it is a widely used and simple environment for testing generosity. While exiting remains informative, we suspect that the stand-alone dictator game may be misleading as a measure of generosity for reasons other than those we have already laid out. For example, one might be tempted to conclude that a dictator who gives nothing does not value fair allocations. However, at least some of our participants reported in post-experiment interviews that they believed keeping all $10 was the action that one *ought* to take and was the action that others expected. One explanation for these beliefs is that teaching students the economic assumption of self-interest leads them to conclude that people are and ought to be self-interested (Frank, Gilovich, & Regan, 1993).

If some people believe that keeping everything is the action they ought to take, the possibility is raised that some people prefer the generous outcome, yet are reluctantly selfish because they think that others expect them not to give. This notion is consistent with Murnighan, Oesch, and Pillutla (2001) notion of managing the "economist" self-impression, and lends to an argument that has been made in different domains by Miller (1999). Miller finds that some people want to be generous, but may need selfish excuses for doing so because they want to comply with a "norm of self-interest." It is difficult to understand fairness and giving—and dictator games purport to help this understanding—if we do not know what people think is expected or fair.

One way to increase consensus would be to construct payoffs such that an equal division is also more efficient, thus capitalizing on all of the motivations identified by well-known social preference models—difference aversion, efficiency, and maximin preferences. For example, if the dictator's choice was between ($6, $1) and ($5, $5), there might be greater consensus around the "fair" outcome than in the standard dictator game. Or, if every dollar given in a standard $10 game was multiplied by 4, we suspect that many, if not most, dictators would be willing to give up at least $2 so that the outcome was a welfare-improving $8 each. At least such games might reduce the number of people who are unsure what the receiver expects of them, compared to standard games in which players are unsure if ($10, $0) is seen as excessively greedy or just smart.

The present results also suggest approaches to modeling fairness and giving. Specifically, the interesting aspects of generosity demonstrated by our studies cannot be adequately formalized without somehow incorporating beliefs into the decision maker's utility function. The literature on psychological games (cf, Geanakoplos et al., 1989) provides a general framework for directly incorporating beliefs and expectations into the analysis of games, and has been applied to strategic contexts like the social dilemma (Rabin, 1993). A similar approach could be useful in understanding basic social preferences in a non-strategic setting. Rather than using some comparison of social payoffs to discount selfish utility, as many current social preference theories do, a comparison of the recipient's expected and realized payoffs may be used.

Battigalli and Dufwenberg (2005) provide a simple utility model for a decision maker that prefers to tip a cab driver whatever amount she thinks the driver expects her to tip, but not more. Applying their idea to the dictator game, the dictator's utility can be given by:

$$U \equiv X - m - \alpha|\mu - m|,$$

where X is the endowment, m is the amount the dictator gives, and μ is the amount that the dictator expects the receiver to expect her to give. The parameter α is here used to represent the individual dictator's sensitivity to the receiver's expectations, which is assumed to be heterogeneous in the population. This model says that if the dictator is sensitive enough, she will conform her behavior to the receiver's expectations. However, she would be better off if her actions were not scrutinized, and may pay a price to lower the receiver's expectations.

Our purpose in presenting this simple model is to suggest a step in the right direction for thinking about generosity. For example, if μ (the expected amount) = $5 in a $10 dictator game, meaning that the dictator thinks the receiver expects half of the endowment, the dictator will give $5 if she cares enough about receiver expectations (e.g., if $\alpha > 1$). Yet, she would prefer that the receiver expected nothing so that she did not have to give. If she could somehow lower the receiver's expectations, she would if the cost of doing so was less than the amount by which expectations were reduced. This behavior is akin to crossing the street to avoid the beggar: We may give to a beggar without really caring about his welfare, since we would just as soon reduce his expectations on us and give him nothing. With this simple framework, we can also see why a dictator who keeps all of a $10 endowment would want to take a $9 exit, as some of our dictators did. It can be that the dictator values appearances (α is positive), but not enough to pay the cost of $5 to be fair. However, if she cares enough to pay $1, she will take the exit option so that the receiver expects nothing. With the utility function above, the dictator will choose to give nothing yet exit if $.2 < \alpha < 1$.

Finally, our results can help in understanding and managing prosocial behavior. For example, we cannot safely assume that people who are generous or charitable genuinely care about others' welfare. In fact, the very people who appear most generous may also be those who are the most motivated to avoid being asked for charity. Similarly, people who behave selfishly may be coerced into behaving more prosocially when put into

situations where expectations are clear and inescapable. We note that many attempts to improve behavior involve inculcating people with the idea that others know or see what they are doing. Children are told that Santa is watching and "knows if you've been bad or good," and public service announcements show Iron Eyes Cody crying when a passing motorist carelessly litters. The latter example is particularly apropos because the motorist is essentially anonymous and the only "punishment" Iron Eyes Cody can levy is his expression of sadness.

We are here reminded of the parable of the Ring of Gyges (related by Glaucon in Plato's *The Republic*). Gyges was a shepherd who found a ring that could turn him invisible. When he used the ring, others did not notice that he disappeared, and simply talked as if he were never there. Gyges, supposedly a man who was good before finding the ring, soon exploited this power to murder the king and usurp the kingdom. The parable was used to argue a "might makes right" perspective on justice, which held that just and unjust people would be indistinguishable if they had Gyges' ring. For Glaucon, Gyges' behavior was normative and not immoral; one who never used this power to selfish ends, it was argued, would be thought of as an idiot while praised aloud for fear of reprisal.

While we do not advocate this view of justice as normative, we find Glaucon's position descriptively insightful. The ultimate might, as demonstrated in the parable, was the ability to control perception; it was not invincibility but invisibility. One who was merely invincible would presumably feel inhibited by being seen and although unpunishable, would be less capable of doing wrong out of self-interest. Gyges' ring, however, allowed him to do misdeeds without others realizing that he, or perhaps anyone, was responsible. In this way, invisibility affords more freedom to be unjust than does total power and anonymity. The dictator game affords the dictator power and anonymity, but not invisibility, since the receiver still knows that *someone* has money to divide. Just knowing that one is the anonymous dictator that the receiver will think badly of can be sufficient to compel giving. Likewise, had Gyges' contemporaries cursed the "someone" who was committing a series of bad acts, Gyges would probably have felt more inhibited knowing he was the someone who the kingdom despised. The exit option in our studies allowed dictators to control perceptions, and that led some to be more selfish than they otherwise would be if left only with power and anonymity.

References

Andreoni Impure altruism and donations to public goods: A theory of warm glow giving. *The Economic Journal, 100,* 464–477.

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

Battigalli, P., Dufwenberg, M. "Dynamic Psychological Games" (April 2005). IGIER Working Paper No. 287. Online: <http://ssrn.com/abstract=707401>.

Bethwaite, J., & Tompkinson, P. (1996). The ultimatum game and non-selfish utility functions. *Journal of Economic Psychology, 17,* 259–271.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes, 63,* 131–144.

Bolton, G., & Ockenfels, A. (2000). A theory of equity reciprocity and competition. *American Economic Review, 100,* 166–193.

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology, 65,* 17–28.

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

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*(3), 817–870.

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior, 30,* 163–182.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47,* 268–298.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives, 3*(4), 99–117.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review, 94,* 857–869.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry, 41,* 20–26.

Fehr, E., & Gachter, S. (2002). Altruistic punishment in humans. *Nature, 415,* 137–140.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition and co-operation. *Quarterly Journal of Economics, 114,* 817–868.

Frank, R., Gilovich, T., & Regan, D. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives, 7,* 159–171.

Gaertner, S. L. (1973). Helping behavior and racial discrimination among Liberals and Conservatives. *Journal of Personality and Social Psychology, 25,* 335–341.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior, 1,* 60–79.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24,* 153–172.

Kahneman, D., Knetsch, J., & Thaler, R. (1986). Fairness and the assumptions of economics. *Journal of Business, 59,* 285–300.

Loewenstein, G., Thompson, L., & Bazerman, M. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology, 57,* 426–441.

Messick, D. M. (1995). Equality, fairness, and social conflict. *Social Justice Research, 8,* 153–173.

Messick, D., & Sentis, K. (1985). Estimating social and nonsocial utility functions from ordinal data. *European Journal of Social Psychology, 21,* 389–399.

Miller, D. (1999). The norm of self-interest. *American Psychologist, 54*(12), 1053–1060.

Murnighan, J. K., Oesch, J. M., & Pillutla, M. (2001). Player types and self impression management in dictator games: Two experiments. *Games and Economic Behavior, 37,* 388–414.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*(5), 1281–1302.