Economic Theory (2007) 33: 67–80
DOI 10.1007/s00199-006-0153-z

SYMPOSIUM

Jason Dana · Roberto A. Weber
Jason Xi Kuang

Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness

Received: 31 August 2005 / Accepted: 18 August 2006 /
Published online: 23 September 2006

Abstract This paper explores whether generosity in experiments is truly evidence of concern for desirable social outcomes. We conduct an experiment using a binary version of the dictator game. We introduce several treatments in which subjects are able to leave the relationship between their actions and resulting outcomes uncertain, either to themselves or to another subject influenced by those actions, thus giving subjects the moral "wiggle room" to behave self-interestedly. We find significantly less generous behavior in these manipulations, relative to a baseline in which the relationship between actions and outcomes is transparent. We conclude that many subjects behave fairly in the baseline case mainly because they intrinsically dislike appearing unfair, either to themselves or others.

We thank Cristina Bicchieri; Iris Bohnet; Colin Camerer; Robyn Dawes; Ernst Fehr; George Loewenstein; John Patty; Charlie Plott; Matthew Rabin; seminar participants at Carnegie Mellon, Berkeley, Cornell, Emory, and Princeton; participants at the 2003 Public Choice, 2003 ESA, 2003 SJDM, and 2004 BDRM meetings; and anonymous referees for helpful comments and suggestions. We greatly appreciate the access to resources at the Pittsburgh Experimental Economics Laboratory (PEEL) at the University of Pittsburgh. This research was funded by a Carnegie Mellon Berkman Faculty Development Grant to Weber and a Russell Sage Behavioral Economics Small Grant to Dana.

J. Dana (✉)
Department of Psychology, University of Pennsylvania, Philadelphia,
PA 19104-6241, USA
E-mail: danajd@psych.upenn.edu

R. A. Weber
Department of Social and Decision Sciences, Carnegie Mellon University,
Pittsburgh, PA 15213, USA
E-mail: rweber@andrew.cmu.edu

J. X. Kuang
College of Management, Georgia Institute of Technology, Atlanta,
GA 30308-0520, USA
E-mail: jason.kuang@mgt.gatech.edu

Keywords Fairness · Experiments · Dictator games

JEL Classification Numbers C91 · D63 · D64

1 Introduction

Subjects across a variety of experiments show apparent concern for others' welfare, beyond any concerns for reputation or punishment. This phenomenon is clearest in dictator games, where a "dictator" makes a one-shot division of an endowment between herself and an anonymous "recipient" who must accept the division. A purely self-interested dictator would keep the entire endowment, but a majority of experimental dictators give a positive amount, and the average amount given is over 20% (see Camerer 2003, chapter 2). Even when elaborate steps are taken to ensure double-blind anonymity, the amount given to an unknown recipient is still often greater than zero (Hoffman et al. 1994).

Several "social preference" theories attempt to capture this apparent generosity by assuming that it reflects a preference for equitable outcomes or social welfare. For instance, people may share with others because they have increasing utility in others' payoffs (Andreoni 1990; Andreoni and Miller 2000), are averse to advantageous payoff differences (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), or want to maximize total social payoffs or the lowest payoff to any one party (Charness and Rabin 2002; Engelmann and Strobel 2004). The common feature of these models is that they assume that dictators' preferences can be characterized by considering only final distributions of wealth. Thus, giving can be interpreted within this framework as being like any other consumption good, except that the dictator is "buying" equity or social welfare.

Yet, other important motives for giving may not be adequately captured with monetary payoffs alone. People may feel compelled to give in some situations—even though they prefer the own-payoff-maximizing outcome—because they do not want to appear selfish, either to themselves or to others. Thus, the underlying motivation driving much fair behavior might be self-interest, coupled with a desire to maintain the illusion of not being selfish. This means that the same people who give in a context like the dictator game may actually prefer the self-regarding and unfair outcome, as long as they have an excuse not to have to give (or be faced with the choice of deciding whether to give).

For instance, a dictator may prefer not to know the consequences of her actions, if possible, in order to not feel compelled to give. Thus, someone on a marrow donor registry, who would give if told she is a match for a recipient, may instead remove herself from the registry to never find out if there is a need to give. Or someone who may have a sexually transmitted disease, and who would feel compelled to stop having unprotected sex were he to know for certain, may avoid testing in order to be able to continue doing so.

Alternatively, people may rely on the presence of other "dictators"—any of whom could help a potential recipient—to not feel compelled to provide help themselves. Thus, in driving by a stranded motorist, one may rely on the possibility that someone else will provide help—even if such help is unlikely—in order not to feel compelled to do so. Or in the case of aiding a crime vi ctim, the presence of many potential helpers may provide everyone with an excuse not to help, thus decreasing the overall likelihood of help being rendered.[1]

Finally, a dictator may exploit (possibly asymmetric) uncertainty about what, precisely, causes unfair outcomes. A dictator who would normally behave generously might, under such "plausible deniability," behave more self-interestedly. For instance, a firm manager may act in a manner that is beneficial for shareholders and employees if it is clear that she is solely responsible for their welfare, but may exploit uncertainty—such as whether poor outcomes are the product of market forces—to behave self-interestedly at their expense.[2]

Considering only preferences over the outcomes associated with giving or not giving, much of the above behavior is self-contradictory. However, these examples are psychologically compelling. Rather than having a preference for a fair outcome, people may conform to situational pressures to give in certain contexts, but may also try to exploit situational justifications for behaving selfishly. Thus, people may appear to act out of concern for others' welfare in some cases, yet behave much more self-interestedly in circumstances that differ only slightly.

We propose a similar mechanism behind much dictator game giving. The standard laboratory dictator game is "transparent"—there is a commonly known one-to-one mapping between the dictator's actions and the outcomes to both parties. For dictators who do not want to appear selfish, this situation can compel generosity. But that does not mean that these dictators prefer a fair outcome as such. Instead, simply removing transparency, as in the examples above, may create the moral "wiggle room" for dictators to behave more selfishly.[3]

We test this proposition with a series of manipulations on an experimental binary dictator game. Each manipulation capitalizes on uncertainty to eliminate transparency, though in each case dictators can still ensure a generous outcome. We find that a majority of dictators are generous in a transparent baseline game. However, our manipulations show that selfishness increases significantly in the absence of transparency.

Our results are important for economic theories of other-regarding behavior. As we mention at the beginning of this paper, the prevalent theoretical approach for understanding sharing in the dictator game has been to introduce utility for fair or welfare-maximizing payoff distributions (Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Charness and Rabin 2002). However, our results cannot be explained by modeling utility only over payoff distributions. If the generosity in the baseline game were motivated by a preference for a certain kind of outcome, then manipulating transparency should be irrelevant as long as dictators can still ensure such outcomes. Instead, we find that decreasing transparency always produces more self-interested behavior.

[1] This is consistent with an extensive literature in social psychology on "bystander intervention" and "diffusion of responsibility" (see Darley and Latane 1968; Latane and Nida 1981).

[2] Indeed, a central part of Enron executives' defense following the firms' collapse was that it did not result form their actions, but instead from external forces such as short sellers and negative press coverage (Stewart 2006).

[3] Moreover, since non-transparency is a common feature of dictator-like situations in naturally occurring economic settings, it is worth exploring what happens to behavior in the laboratory when this feature is introduced (see Weber and Camerer 2006).

Of course, this paper is not the first to point out problems with models that account for fair behavior solely by relying on preferences over payoff distributions. For example, reciprocity and perceptions of others' intentions can be important in determining one's utility for a social outcome (e.g. Rabin 1993; Dufwenberg and Kirchsteiger 2004; see also, Blount 1995; Falk et al. 2003; Bolton and Ockenfels 2005). However, reciprocity and others' intentions are inconsequential in our setting because the receiver is passive. Thus, our results demonstrate that more than distributional concerns drive fair behavior even in settings involving unilateral action.[4]

2 Experimental design

Subjects took part in a modified dictator game with a binary choice between an equal and an unequal (and welfare inefficient) wealth allocation. The baseline game was transparent in the manner of standard dictator games, while three subsequent manipulations relaxed transparency in the manner of the examples we discussed previously. Subjects were randomly assigned to only one treatment.

2.1 General procedures

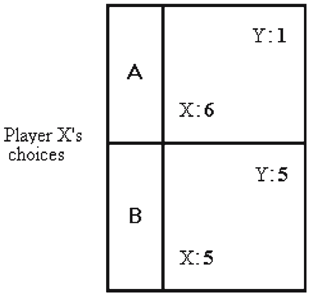
Subjects were undergraduates at the University of Pittsburgh, who participated voluntarily in response to advertising for paid decision experiments. All experimental sessions were run with at least 12 subjects present. Upon arriving at the experiment, subjects were seated at computer terminals, through which they received instructions that were also read aloud.[5] All experimental stimuli were presented via computer interface, and all interaction occurred via the computers.

At the beginning of the experiment, subjects drew cards to determine their role (letter) and matching (number). Subjects were instructed that they would be playing a simple game with one other person in the room (or two other people, in the multiple dictator treatment), with whom they were matched anonymously and randomly. Subjects were told that all members of their group would be paid according to the choice made by the dictators (Player "X" in most treatments, Players "X" and "Y" in the multiple-dictator treatment).

After receiving instructions describing a generic payoff table, subjects completed a short quiz to ensure that the task and the payoff representation were understood. Subjects were then shown the actual payoffs for the experiment and any other necessary information to describe their particular treatment. We conducted four treatments, using a total of 190 subjects: *baseline* (38 subjects, 19 dictators), *hidden information* (64 subjects, 32 dictators), *multiple dictator* (30 subjects, 20

[4] Our work is more closely related to that of Rabin (1995), Bolton et al. (1998), and Konow (2000), who posit that fair behavior is driven by comparisons against a standard, but that such a standard serves mainly as a constraint that individuals seek to circumvent rather than a goal that they seek to implement.

[5] Instructions are available at http://www.psych.upenn.edu/∼dana.



Player X's choices

A	Y:1 X:6
B	Y:5 X:5

Fig. 1 Interface for baseline treatment

dictators), and *plausible deniability* (58 subjects, 29 dictators). We now describe each treatment in detail.

2.2 Baseline

The baseline payoffs, as shown to subjects, are presented in Fig. 1. Dictators ("Player X") chose between *A* and *B* by clicking on one of the two letters. Before the software allowed them to make a choice, subjects were given 60s—during which the payoff matrix and choice interface were displayed on the screen—to consider what they would do. When dictators made their choices, receivers were asked to choose hypothetically between the two options, serving in part to maintain the anonymity of the roles. Upon completion of the game, subjects were paid privately as they exited the room.

In the baseline treatment, the relationship between actions and outcomes is transparent. We also conducted three manipulations that eliminated this transparency. The procedures for each of these treatments were identical to the baseline unless otherwise noted.

2.3 Hidden information treatment

In this treatment, we allowed a dictator to remain ignorant to the precise consequences to the recipient. Each dictator (again "Player X") again received \$6 for choosing *A* and \$5 for choosing *B*, but the receiver's payoffs from these actions were uncertain. Subjects were informed that the receiver's payoffs from *A* and *B* were determined by a coin flip prior to the session and could have been \$1 and \$5, respectively (as in the baseline), or "flipped" (\$5 and \$1, respectively) so that choosing *A* made both parties better off.

We conducted four sessions, two for each set of payoffs, and a total of 16 dictators were assigned to each payoff set. Subjects were instructed that the true payoffs would not be revealed publicly, but that Player X could reveal them by clicking a button. All subjects were informed that Player X's decision of whether to reveal would be kept private from Player Y. The matrices representing this game to subjects are presented in Fig. 2. Subjects were told that clicking the reveal button would replace the question marks with the receiver's true payoffs. When Player Y

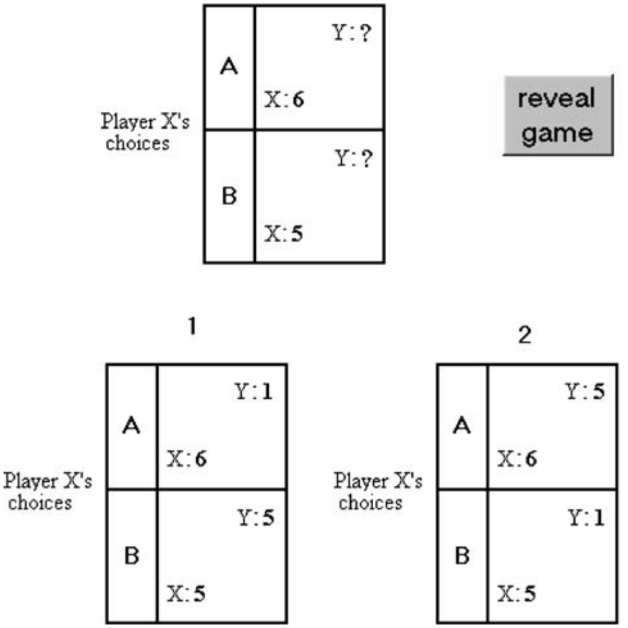


Fig. 2 Interface for hidden information treatment

subjects made their hypothetical choices, they were asked to assume the conflicting interest payoffs (i.e., the same payoffs as in the baseline; matrix 1 of Fig. 2).

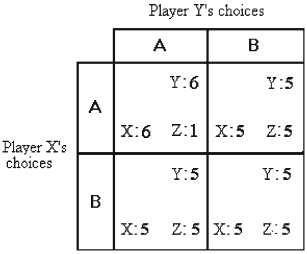
If giving in the transparent baseline game reflects a preference for an equitable payoff distribution, then the proportion of dictators who give in the baseline should be equal to the proportion that reveals the true payoffs and chooses the most equitable action in the hidden payoff treatment. Instead, if dictators are seeking an excuse to not feel compelled to give, then we might expect them to choose to remain uninformed and to choose A under ignorance.

2.4 Multiple dictator treatment

In this treatment, we added a second dictator to the baseline game, thus eliminating each dictator's sole responsibility for the unfair outcome. However, either dictator could independently implement the fair outcome.

Two experimental sessions were run, each with 15 subjects. Two-thirds of the subjects (20) were assigned to strategic player roles ("Players X and Y"), the rest were passive recipients ("Player Z"). Subjects were informed that all three players would be paid according to the combined choices of Players X and Y, as depicted in the matrix in Fig. 3. While subjects assigned to the role of X or Y made their choices, those assigned to the role of Z indicated which option they thought the majority of players would choose.

Because both dictators must choose *A* to obtain the inequitable outcome ($6, $6, $1), the addition of a second dictator does nothing to impede subjects from ensuring a fair outcome if they prefer. Either dictator can impose the fair outcome ($5, $5, $5) by choosing *B*. Thus, we might predict the same proportion of *B* choices in this



		Player Y's choices	
		A	B
Player X's choices	A	Y:6 X:6 Z:1	Y:5 X:5 Z:5
	B	Y:5 X:5 Z:5	Y:5 X:5 Z:5

Fig. 3 Interface for multiple dictator treatment

treatment as in the baseline.[6] However, this treatment breaks transparency—the selfish choice *A* no longer guarantees the unfair outcome for the passive recipient. Thus, by choosing *A*, a strategic player can allow the selfish option to result, while not having implemented it directly.

2.5 Plausible deniability treatment

In the final treatment, we allowed a dictator the (unlikely) possibility of losing agency, thus allowing outcomes to plausibly result from causes other than the dictator's actions.

Three experimental sessions were run with at least 18 subjects at each. A "cutoff" feature was added to the baseline game. Subjects were informed they would have a 10s interval during which to enter their choices, but that if they had not already chosen at a randomly selected point in the interval, the software would cut them off and choose between *A* and *B* with equal probability. Subjects did not know the precise cutoff point, only that it could occur anywhere in the 10 s interval. Only the dictator would be notified if a cutoff occurred, so that receivers could not be sure if their payoffs were determined by the dictator's choice or the software. The choice interval took place after subjects were given one minute to consider their choices—during which the choice interface appeared on-screen as in earlier treatments—and did not begin until subjects clicked a "begin game" button. Receivers' hypothetical choices were also subject to the possibility of a cutoff. Subjects who were cut off were asked to indicate how they would have chosen had they not been cut off.

The cutoff points were drawn from a discretized normal distribution.[7] Our goal was to allow dictators a reasonable amount of time to choose before being cut off.

[6] Of course, there is a difference between the two treatments in that a choice of B now affects the other strategic player who may have preferred the inequitable outcome. However, the payoff difference for the other player is 1, while this difference is 4 for the receiver, implying that a subject in the role of Player X or Y would have to care about the other strategic player four times as much as she cares for the receiver in order for this difference to completely compensate for differences in equity between the two outcomes. If, as is more likely the case, the welfare of both other "players" is valued equally, then the loss of one should only matter slightly. Moreover, this loss only negatively affects other players who preferred outcome A to B in the baseline condition treatment, which we saw were a minority.

[7] To be precise, cutoffs occurred exactly on one of the 10s, with the greatest mass at 5 and 6s. The mass placed on seconds 5 and 6 was equivalent to the area in the first standard unit (0.34), seconds 4 and 7 the second standard unit (about 0.13), etc.

Pre-testing revealed that when people were asked to choose quickly after clicking the “begin game” button, they were always able to do so in less than 2 s. The probability of being cut off in less than 2 s was about 3×10^{-5}, and no one was actually cut off in less than 4s, meaning that a dictator truly interested in making a choice would have enough time to do so. Thus, the cutoff feature should be largely irrelevant if dictators’ behavior is driven by preferences over final payoff distributions.

However, this feature relaxes transparency and also allows us to discriminate between two distinct possible mechanisms for moral wiggling. First, because receivers are never able to differentiate between dictators and nature in how their payoffs are determined, dictators could feel free to choose *A* more frequently, an *other-deceptive* motive. However, dictators would still know that they were responsible for ensuring an inequitable outcome. Thus, if a *self-deceptive* motive is instead responsible for moral wiggling, then perhaps dictators will “dither” and allow themselves to be cutoff. With half-probability, the software would choose the fair outcome they would have felt compelled to choose anyway, but with half-probability the selfish outcome would obtain and the dictator could maintain the illusion of not being responsible for its implementation.[8]

3 Results

3.1 Baseline game

As expected, a majority of dictators acted fairly. Of the 19 dictators, 14 (74%) chose *B*, the ($5, $5) option. Further, all 19 receivers hypothetically chose *B*.

The sort of generosity seen in our baseline game is consistent with previous evidence of sharing in dictator games (see Camerer 2003, chapter 2; Kahneman et al. 1986) and is typically interpreted as supporting the idea that people prefer the generous outcome. However, it is also consistent with the idea that dictators feel compelled to give in transparent situations, of which this is a case. Results from our non-transparent treatments help separate these motives.

3.2 Hidden payoff treatment

The results are summarized in Table 1, which shows the proportion of subjects making each choice, depending on the true underlying payoffs, and the corresponding proportion in the baseline. Of the 16 dictators who faced the same payoffs as in the baseline, ten (63%) chose *A*, resulting in the ($6, $1) outcome. This behavior resulted in spite of the fact that dictators could costlessly reveal that the payoffs were exactly the same as in the baseline treatment, where a majority of dictators (74%) chose *B*. The difference in these proportions is statistically significant [$\chi^2(1) = 4.64$, $p = 0.03$].[9] Moreover, as Table 1 also reveals, only 18 of all 32

[8] Allowing one’s self to be cut off is also interesting behavior because it implies preferring a mixture of two outcomes over each one separately, which is inconsistent with a theory of rational choice with utilities defined only over outcomes.

[9] This result has been replicated using double-blind anonymity and forcing dictators to take an action (rather than just remain passive) if they wish to remain ignorant (Larson 2005), as well as using a within-subjects design with varying probabilities of the two payoff states (Munyan 2005).

Table 1 Comparison of baseline and hidden information treatments

Treatment	Proportion choosing "A" (unfair choice)	Proportion revealing true payoffs
Dictators' (Player X) choices		
Baseline	5/19 (26%)	
Hidden information (Matrix 1, baseline payoffs)	10/16 (63%)	8/16 (50%)
Hidden information (Matrix 2, alternate payoffs)	13/16 (81%)	10/16 (63%)
Recipients' (Player Y) hypothetical choices		
Baseline	0/19 (0%)	
Hidden information (Matrix 1, baseline payoffs)	13/32 (41%)	

Table 2 Allocation choices by information acquisition in hidden information treatment

Actual payoffs	Information acquisition choice	Proportion choosing "A"
Matrix 1 (baseline payoffs)	Chose to reveal (8/16, 50%)	2/8 (25%)
	Chose not to reveal (8/16, 50%)	8/8 (100%)
Matrix 2 (alternate payoffs)	Chose to reveal (10/16, 63%)	9/10 (90%)
	Chose not to reveal (6/16, 38%)	4/6 (67%)

dictators (56%) chose to reveal the true payoffs. Thus, behavior differs significantly from the baseline—even with identical payoffs—and many subjects' behave consistently with a desire to remain ignorant to the consequences of a self-interested choice.

Table 2 presents choices broken down by true underlying payoffs and information acquisition behavior. As we discuss above, dictators motivated by a preference for socially desirable outcomes should reveal the true payoffs and act fairly. Therefore, we would expect the proportion of dictators acquiring information and choosing the option that gives \$5 to the recipient to be comparable to the proportion choosing *B* in the baseline (74%). However, as Table 2 shows, only 15 of 32 dictators (47%) revealed the true state *and* chose the other-regarding option, considerably less than the proportion of generous dictators in the baseline game [$\chi^2(1) = 3.49$, $p = 0.06$].[10]

Receivers' hypothetical choices mirrored those of the dictators. While all receivers stated they would choose the fair option in the baseline treatment, only 59% said they would so in the hidden payoff treatment [$\chi^2(1) = 10.36$, $p = 0.001$].[11]

The hidden payoff results suggest that several of the generous choices in the baseline were not the result of dictators wanting to implement the (\$5, \$5) outcome. Instead, many dictators appear to exploit the payoff uncertainty as an excuse for

[10] In fact, this proportion likely *overestimates* the frequency of other-regarding behavior. Choosing *A* with matrix 2 is *both* selfish (it yields the highest payoff for the sender) and other-regarding (it yields the highest payoff for the receiver). Thus, any dictator who looked, found matrix 2, and chose *A* would count as choosing "fairly"—even though such behavior is consistent with purely self-interested motives.

[11] One could argue that without incentives, the receivers are exhibiting a socially desirable response bias. Nevertheless, the fact that receivers' hypothetical behavior changes in the same manner as the behavior of dictators suggests agreement regarding the appropriateness of the two choices in the two environments.

Table 3 Choices by dictators in baseline and multiple dictator treatments

	Proportion choosing "A"
Multiple dictator	13/20 (65%)
Baseline	5/19 (26%)

behaving self-interestedly. In the multiple dictator treatment, we explore whether this pattern obtains when dictators can ensure the fair outcome, but choosing self-interestedly does not ensure inequity.

3.3 Multiple dictator treatment

The multiple dictator and baseline treatments are compared in Table 3. While 74% of subjects chose fairly (*B*) in the baseline, only 35% did so in the case with two dictators ($\chi^2(1) = 5.87$, $p = 0.02$).[12] Further, it seems that receivers (Player Z) shared our intuition. All ten correctly predicted that *A* would be the most common choice by strategic players, in contrast with receivers' hypothetical choices in the baseline game, where no one chose *A*.[13]

These results further strengthen our contention that a great deal of giving does not result from a desire to implement socially desirable outcomes.[14] In this treatment, the option of ensuring equity and maximizing welfare was always available to strategic players, but it was not transparent that choosing self-interestedly (*A*) would produce an unfair outcome.

3.4 Plausible deniability treatment

In both modifications thus far, eliminating transparency resulted in decreased giving, relative to the baseline. The results of the plausible deniability treatment allow us to examine whether such moral wiggling is produced by dictators capitalizing on receivers' incomplete information (i.e., "other-deception") or by their engaging in "self-deceptive" reasoning.

Table 4 lists the choice proportions and the proportion of subjects who were cut off. Among dictators who were not cut off (22/29), a majority (55%) chose

[12] The equitable outcome occurred in five of ten groups (the expected proportion, based on actual choices, is 59%). This is lower than the proportion of fair choices in the baseline (74%), but this difference is not significant.

[13] Of course, this comparison is slightly awkward since in the baseline the recipients indicate their own hypothetical action while in this treatment they indicate their expectation of the behavior of the other players. We make the comparison simply to illustrate that the expectations of the passive participants about appropriate/actual behavior show a similar pattern to the actual behavior of dictators.

[14] Alternatively, choosing *A* in the multiple dictator treatment could be interpreted as a "you decide" option for a subject uncertain about which action is more socially desirable or the appropriate norm. This is consistent with our interpretation, in which both dictators, neither of whom wants to act generously, use the excuse of leaving it up to the other to do so. Thus, exploiting such "you decide" norms might be a mechanism through which moral wiggling occurs, a hypothesis consistent with the observation that the change in behavior in Table 3 is larger than that usually observed solely as a result of such norms acting alone (cf. Cooper and Van Huyck 2003).

Table 4 Results in the plausible deniability treatment

	Dictators ($n = 29$)	Receivers hypothetical choices ($n = 29$)
Proportion cutoff	7/29 (24%)	11/29 (38%)
Average cutoff time for those cutoff	4.30	4.64
Proportion of A choices by those not cutoff	12/22 (55%)	5/18 (28%)
Total number of A outcomes	17/29 (59%)	
Proportion of those cutoff stating they would have chosen A	1/7 (14%)	3/11 (27%)

the selfish action *A*, a higher proportion than chose *A* in the baseline [$\chi^2(1) = 3.35, p = 0.07$]. Thus, receivers' uncertainty about how payoffs are determined appears sufficient to promote increased self-interest, even for dictators who know the consequences of their actions.

However, a substantial proportion of dictators (24%) was cut off and did not make a choice. The average cutoff time for these dictators was approximately 4.3s, with none of the cutoffs occurring before 4s. (Recall that no one took longer than 2 s to make a choice in a pretest.) Thus, it appears that many subjects were willing to delay making a choice, with the hope of avoiding making a choice altogether.

Overall, only 10 out of 29 choices (34%) are consistent with a desire to implement the fair outcome.[15] However, dictators engaging in moral wiggling are heterogeneous in how they ultimately obtain selfish outcomes—while some dictators directly choose *A* (exploiting the recipients' uncertainty), others allow themselves to be cutoff by the computer (exploiting their own lack of agency and uncertainty over outcomes).

Moreover, similarly to previous treatments, we again see that only a minority of recipients' hypothetical choices, 13 of 29 (45%) are consistent with implementing the fair outcome [in contrast to every recipient in the baseline; $\chi^2(1) = 15.72, p = 0.001$]. Thus, recipients' intuitions again capture the effect of our treatment on behavior.

4 Conclusions

Generosity in experiments is often interpreted in behavioral economic models as a preference for a fair or efficient outcome. However, our experiments indicate that a good deal of giving is consistent with another interpretation: people feeling compelled to give due to situational factors, while not really valuing the corresponding outcome. To test this possibility, we relaxed the transparency—common knowledge of a one-to-one mapping between actions and outcomes—of standard dictator experiments. We find that relaxing this property, thus giving dictators the

[15] Even if we assume that those dictators who got cut off were equally likely to choose either A or B (for instance, if they were indifferent between the two options) the results do not change substantially. Assigning half of these dictators to each choice yields a majority of A choices (15.5/29, or 53%). This is still greater than the proportion of A choices in the baseline [$\chi^2(1) = 3.45, p = 0.06$].

Table 5 Proportion of dictators implementing fair outcome across treatments

Treatment	Proportion implementing fair outcome
Baseline	14/19 (74%)
Hidden information (baseline payoffs)	6/16 (38%)
Multiple dictator	7/20 (35%)
Plausible deniability	10/29 (34%)

moral "wiggle room" to behave self-interestedly while maintaining the illusion of fairness, significantly decreases fair behavior.

Moreover, as Table 5 reveals, there is a striking regularity across our modified treatments. In *all* three treatments that relax transparency, the proportion of subjects implementing the fair outcome—which subjects could do in all treatments—is roughly 35%, or half of what it is in the baseline.

Thus, the overall pattern of our results is consistent with three kinds of motivations for other-regarding behavior. Roughly one-third of subjects act as though they value implementing fair outcomes, while about a quarter behave self-interestedly even in a transparent situation such as the baseline (see Table 5). The differences in proportions between the baseline and non-transparent treatments, however, reflects behavior that is contextually driven, in some instances reflecting an apparent preference for fair outcomes, but in others self-interest. Importantly, this pattern is quite similar to that found in other recent experiments measuring social preferences under the possibility of moral wiggling (cf. Dana et al. 2006; Lazear et al. 2006).

One way to view our aggregate results is that there are environments with strong prescriptions for fair behavior (such as feeling compelled to act generously in transparent dictator games), but that these norms or constraints are less binding, or perhaps compete with other norms (see footnote 14), once transparency is eliminated. In such cases, we conjecture that the norms most consistent with self-interest will often exert stronger influences on behavior. For instance, in settings corresponding to the hidden information treatment, dictators might rely on the "mind your own business" norm to justify not acquiring information on the other party's payoff.

Two important caveats must accompany our results. First, a significant amount of sharing *is* consistent with the idea that people value implementing fair outcomes. Across treatments, roughly one-third of subjects do so (see Table 5), even when non-transparency leads others not to. Second, our main result is *not* that subjects choose to act self-interestedly at the expense of another. Instead, we claim that it is precisely the lack of certainty regarding the consequences to the other party that allows much of the self-interested behavior in our modified treatments to occur. In fact, the proportion of dictators who choose—with certainty—to implement the unfair outcome is never greater than one-half.[16]

We should also note that our work overlaps with other existing research on the determinants of fair behavior. For instance, previous studies demonstrate that people capitalize on uncertainty or information asymmetries to behave more

[16] The proportions are: 24% (baseline), 13% (hidden information), and 41% (plausible deniability). It is impossible for a dictator to implement the unfair outcome alone in the multiple dictator treatment.

self-interestedly, though usually to avoid sanctions or only to keep others ignorant of whether the outcomes obtained are fair (Roth and Murnighan 1982; Mitzkewitz and Nagel 1993; Kagel et al. 1996; Dana et al. 2006; Lazear et al. 2006). Similarly, other studies explore the relationship between fair behavior and self-impressions, arguing as we do that much fair behavior is the product of a desire to maintain positive identifications (Murnighan et al. 1999; Akerlof and Kranton 2000; Benabou and Tirole 2005). In addition, a significant body of work explores external social norms—regarding what actions are considered socially appropriate—and how these influence behavior (Cialdini et al. 1990; Bicchieri 2006; Shang and Croson 2005; Krupka and Weber 2006).

As with our work, a key goal of all this research is to improve our understanding of the determinants of fair behavior, with the ultimate goal of improving economic theory. Our contribution is to show that a key feature of many experiments measuring fairness, transparency, plays an important role in the choices people make and kinds of outcomes that result, and does so in a systematic manner.[17] Of course, a broader goal is that our work, combined with the other kinds of research described above, will lead to an improved theoretical understanding of what drives fair behavior.

References

Akerlof, G., Kranton, R.: Economics and identity. Q. J. Econ. **115**(3), 715–753 (2000)

Andreoni, J.: Impure altruism and donations to public goods: a theory of warm glow giving. Econ J **100**, 464–477 (1990)

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

Benabou, R., Tirole, J.: Incentives and prosocial behavior (2005) (unpublished manuscript)

Bicchieri, C.: The Grammar of Society: the Nature and Dynamics of Social Norms. Cambridge, MA: Cambridge University Press (2006)

Blount, S.: When social outcomes aren't fair: the effect of causal attributions on preferences. Organ Behav Hum Decis Process **63**(2), 131–144 (1995)

Bolton, G.E., Katok E., Zwick, R.: Dictator game giving: rules of fairness versus acts of kindness. Int J Game Theory **27**(2), 269–299 (1998)

Bolton, G.E., Ockenfels, A.: A theory of equity reciprocity and competition. Am Econ Rev **100**, 166–193 (2000)

Bolton, G.E., Ockenfels, A.: A stress test of fairness measures in models of social utility. Econ Theory **25**(4):957–982 (2005)

Camerer, C.: Behavioral Game Theory: experiments on Strategic Interaction. Princeton, NJ: Princeton University Press (2003)

Charness, G., Rabin, M.: Understanding social preferences with simple tests. Q J Econ **117**, 817–869 (2002)

Cialdini, R.B., Reno, R.R., Kallgren C.A.: A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. J Personal Soc Psychol **58**, 1015–1026 (1990)

[17] Of course, there are many possible extensions that could further clarify the precise mechanisms driving moral wiggling, that we leave for future work. For instance, in the multiple dictator treatment, one could make it so that either dictator acting self-interestedly could implement the inequitable outcome. Alternatively, in the plausible deniability treatment, one could have dictators explicitly choose between making the choice themselves or via a randomizing device (perhaps allowing dictators to manipulate the probability of each outcome). In these cases, we anticipate that at least some of the difference with the baseline will be attenuated due to diminished opportunity for self-deception. We also note that others have already explored some extensions in the context of the hidden information treatment (Larson 2005; Munyan 2005).

Cooper, D., Van Huyck, J.: Evidence on the equivalence of the strategic and extensive form representation of games. J Econ Theory **110**(2), 290–308 (2003)

Dana, J., Cain, D.M., Dawes, R.M.: What you don't know won't hurt me: costly (but quiet) exit in a dictator game. Organ Behav Hum Decis Process **100**(2), 193–201 (2006)

Darley, J., Latane B.: Bystander intervention in emergencies: diffusion of responsibility. J Personal Soc Psychol **8**, 377–383 (1968)

Dufwenberg, M., Kirchsteiger G.: A theory of sequential reciprocity. Games Econ Behav **47**, 268–298 (2004)

Engelmann, D., Strobel, M.: Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. Am Econ Rev **94**, 857–869 (2004)

Falk, A., Fehr E., Fischbacher, U.: On the nature of fair behavior. Econ Inq **41**(1), 20–26 (2003)

Fehr, E., Schmidt, K.M.: A theory of fairness, competition and cooperation. Q J Econ **114**, 817–868 (1999)

Hoffman, E., McCabe, K., Shachat, K., Smith, V.: Preferences, property rights and anonymity in bargaining games. Games Econ Behav **7**, 346–380 (1994)

Kagel, J., Kim, C., Moser, D.: Fairness in ultimatum games with asymmetric information and asymmetric payoffs. Games Econ Behav **13**, 100–110 (1996)

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

Konow, J.: Fair shares: accountability and cognitive dissonance in allocation decisions. Am Econ Rev **90**, 1072–1091 (2000)

Krupka, E., Weber, R.A.: The focusing and informational influences of norms on pro-social behavior (2006) (unpublished manuscript)

Larson, T.: The use of strategic ignorance in dictator games when payoffs are not transparent: computer vs. paper testing techniques. Department of Economics Master's Thesis, Emory University (2005)

Latane, B., Nida, S.: Ten years of research on group size and helping. Psychol Bull **89**(2), 308–324 (1981)

Lazear, E., Malmendier, U., Weber, R.A.: Sorting opportunities in decisions involving fairness (2006) (unpublished manuscript)

Mitzkewitz, M., Rosemarie N.: Experimental results on ultimatum games with incomplete information. Int J Game Theory **22**(2), 171–198 (1993)

Munyan, L.: Patterns of information avoidance in binary choice dictator games. Working paper (2005)

Murnighan, J.K., Oesch, J.M., Pillutla, M.: Player types and self-impression management in dictator games: two experiments. Games Econ Behav **37**, 388–414 (1999)

Rabin, M.: Incorporating fairness into game theory and economics. Am Econ Rev **83**, 1281–1302 (1993)

Rabin, M.: Moral preferences, moral constraints, and self-serving biases (1995) (unpublished manuscript)

Roth, A.E., Murnighan, J.K.: The role of information in bargaining: an experimental study. Econometrica **50**(5), 1123–1142 (1982)

Shang, J., Croson, R.: Field experiments in charitable contribution: the impact of social influence on the voluntary provision of public goods (2005) (unpublished manuscript)

Stewart, J.B.: Common sense: Enron defense wins the award for year's worst. Wall Str J (East Edn) May 31, D.3 (2006)

Weber, R.A., Camerer, C.F.: Behavioral' experiments in economics. Exp Econ **9**(3), 281–295 (2006)