# Strategic Ignorance and the Robustness of Social Preferences

Zachary Grossman*

August 12, 2010

**Abstract**

How robust are social preferences to variations in the environment in which a decision is made? By varying the elicitation method and default choice in the 'moral wiggle-room' game of Dana, Weber, and Kuang (2007), I examine the robustness and nature of the pattern of information avoidance in which many dictators in experiments—if initially uncertain—avoid learning whether their choice will help or hurt another person and choose selfishly. When ignorance is not the default choice, participants choose it much less frequently. However, when dictators express their outcome choice using the strategy method, most are willing to overcome the default choice and reveal the payoff state *ex post*. I conclude that people will employ strategic ignorance to avoid a morally-fraught decision if they can do so passively, but having to actively choose ignorance betrays its usefulness and leads to behavior largely consistent with models of preferences over outcomes. Thus while opportunities to create and exploit moral wiggle-room limit fair-minded behavior, environmental or psychological variables may reinforce the motivation that leads people to choose fair outcomes.

**Keywords:** social preferences, strategic ignorance, moral wiggle-room, default effects, status quo bias, self-deception, self-signaling, dictator games

**JEL codes:** C91, D03

*UC Santa Barbara. Email: grossman@econ.ucsb.edu. I am grateful to Gary Charness, Aldo Rustichini, and Jeroen van de Ven for helpful comments.

# 1 Introduction

Decades of research have documented the willingness of experiment participants to give up money in order to help anonymous recipients.[1] How robust are these social preferences, though? More recent research calls into question the importance of social-preferences in environments in which the decision-maker has the freedom to manipulate her own choice set, the information that she has about her choice, or the information that others have—as is the case in a vast number of the allocation decisions likely encountered by individuals navigating their way through their natural social environment.[2] By documenting giving behavior inconsistent with preferences defined solely over outcomes, this sub-literature offers an important critique of the broader social-preferences literature, which had previously focused on models of preferences over the distribution of material payoffs and on reciprocity.[3] This view of moral behavior as fickle and flickering has been embraced by many as evidence that other, more belief-dependent psychological motivations, such as concern for self- or social-image, guilt, shame, blame, and responsibility can better account for observed patterns of giving.[4]

However, some important questions remain. How important are the forces that undermine the expression of social preferences? To what extent do environmental or psychological variables actually reinforce, as opposed to suppress, the motivation that leads people to choose fair outcomes? This paper presents an experiment that contributes towards answering these questions. It uses the moral wiggle-room game of Dana, Weber, and Kuang (2007) (hereafter DWK), which demonstrated that many dictators—if initially uncertain—will avoid costless information about how their choice affects the payoff of the recipient and, as a result, significantly fewer dictators take the simple steps necessary to guarantee the fair outcome than do in the full-information baseline game. Beyond simply behaving more selfishly when

[1] Camerer (2003) synthesizes the 'first generation' of social-preference models.

[2] See, for example, Dana, Weber, and Kuang (2007), Lazear, Malmendier, and Weber (2009), Dana, Cain, and Dawes (2006), and Broberg, Ellingsen, and Johannesson (2007).

[3] See for example Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002); Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006).

[4] See, for example, Benabou and Tirole (2010), Benabou and Tirole (2006), Andreoni and Bernheim (2009), Charness and Dufwenberg (2006), and Grossman (2009b).

the link between actions and outcomes is obscured, the DWK participants actually seek to obscure this link using *strategic ignorance*, the purposeful avoidance of (free) payoff-relevant information in pursuit of some objective.

The experiment replicates the moral wiggle-room game, then adapts it by varying the default information choice, specifically, whether the dictator's decision whether to reveal the recipient's payoffs is an active choice with no default or whether the default choice is to reveal. In a separate condition, the dictator states her outcome choice using the strategy method and chooses whether to overcome the default choice of not revealing. This allows me to ask: 1) is the ignorance observed by DWK really *strategic* or is it prevalent for other reasons; and 2) what purpose does strategic ignorance serve; are participants just as willing to avoid information after they have indicated their choice as they are beforehand?

Answering these questions sheds light on the ease with which individuals can create and exploit "moral wiggle-room" to pursue self-interest and thus, the external validity of laboratory experiments documenting behavior consistent with a preference for fair outcomes. Furthermore, a deeper understanding of strategic ignorance is necessary to gain insight into the motivation and the cognitive processes that lead people to give in transparent decisions, yet avoid transparency.

While the DWK findings of strategic ignorance and behavior inconsistent with a preference for fair outcomes are robust in the replication, I find that they are quite sensitive to changes in the choice elicitation method. Switching the default choice to reveal, or even requiring an active choice drastically decreases the rate at which participants avoid learning the recipients' payoffs. As a result, behavior in these variants of the moral wiggle-room game *is* consistent with a preference for fair outcomes.

Such large default effects are consistent with the notion that there is some activation cost to overcoming the default and—taken alone—would cast doubt on the notion that information avoidance is strategic. However, the high rate at which participants actively choose to reveal the true payoffs when they have made conditional choices using the strategy method suggests that strategic ignorance is not "just" a default effect. Indeed, the desire

to quickly resolve uncertainty *ex post* drives most participants to overcome whatever bias favors the default.

Thus, people decline information in order to avoid facing an informed decision, not to avoid learning the state. However, having to actively choose ignorance neutralizes the strategic benefit of avoiding the informed decision, so behavior is inconsistent with baseline giving only when ignorance is the default choice. This finding suggests that while opportunities to create and exploit moral wiggle-room limit fair-minded behavior, environmental or psychological variables may reinforce the motivation that leads people to choose fair outcomes, either by making it difficult for people to avoid feeling responsible for the well-being of others or by drawing attention to one's own behavior.

## 2 Experimental Design

The experiments took place at the Experimental and Behavioral Economics Laboratory (EBEL) at the University of California, Santa Barbara. Participants were randomly recruited from the EBEL subject pool (largely comprised of UCSB students and staff) using the online system ORSEE (Greiner 2003). Upon arriving at the experiment, participants sat at computer terminals, through which they read the instructions and entered their decisions. The interface was programmed using the Z-Tree software package (Fischbacher 2007). Screenshots of the instructions as well as the decision interfaces for each condition are presented in the Appendix.

Participants played the same binary dictator game used by DWK. They were instructed that they would be playing a simple game with one other person in the room with whom they had been randomly and anonymously matched, with both members of the group being paid according to the choice of the dictator. The dictator was referred to as 'Player X' and the recipient as 'Player Y'. After participants read instructions describing a generic payoff table, they completed a short quiz to ensure that the task was understood. Next they were shown the actual payoffs for the experiment and any other information relevant to their particular

experimental condition, before taking another short quiz. Before participants were told to which role they had been assigned and were allowed to make a choice, they were given sixty seconds—during which the payoff matrix or matrices were displayed on the screen—to consider their choice. In general, the screen progression and layout reproduced the DWK interface as faithfully as possible. The text of the general instructions were reproduced almost verbatim, as were the condition-specific instructions in the replication conditions.[5]

The dictator always received \$6 for choosing $A$ and \$5 for choosing $B$, but there were two possibilities for the recipient's payoffs. In the 'conflicting interests' ($CI$) version of the game the recipient's payoffs from $A$ and $B$ were \$1 and \$5, respectively, while in the 'aligned interests' ($AI$) version the recipient's payoffs were flipped—\$5 and \$1, respectively. As in DWK, the *baseline* condition featured a transparent dictator game featuring the $CI$ payoffs, while the remaining conditions relaxed this transparency by having the dictator initially unsure of the version, having been told that each was equally likely.

While the dictators made their choices, recipients were asked to choose hypothetically between the two options for each of the two versions of the game, with the exception that in the baseline condition they were only asked about the $CI$ version. Furthermore, participants were asked about their beliefs. All recipients were asked to state their beliefs about the percent of dictators in their session (in relevant conditions: who were knowingly) playing the conflicting-interests game that chose option B and in the relevant conditions, the percent of dictators in the session who chose to reveal the payoffs. Dictators were asked to state their beliefs about the mean of the responses of the recipients in the same session.[6] Participants whose responses were within five percentage points of the correct answer were paid an additional five dollars.[7] Upon completion of the experiment, participants were paid privately in cash as they exited the room.

[5] I am grateful to Jason Dana for sharing the original program used to conduct the experiment. Minor differences in layout arose because the DWK experiment was programmed using a different software package.

[6] Results of these decisions are not central to the questions and conclusions of this paper and are available from the author upon request.

[7] This elicitation method was used in Charness and Dufwenberg (2006) and has the advantages of simplicity and immunity to risk aversion.

The five experimental conditions are described below. The first two exactly replicate the DWK experiment while three additional conditions feature variations of the choice elicitation method. The *Default NR*, *Active Choice*, and *Default R* conditions are designed to provide insight into the role of the default choice in determining the dictator's choice of information. The *Strategy Method* condition highlights the dictator's information choice when she cannot avoid stating her preference in *CI* by remaining ignorant.

1. *Baseline*: This condition exactly replicated the DWK baseline condition. Dictators played the 'conflicting interests' game with certainty. Recipients made a hypothetical choice for the same game. The link between actions and outcomes was transparent. Then the participants answered the role-specific beliefs-elicitation question.
2. *Default NR*: This condition replicated the "hidden information" condition of DWK. The participants were presented with the two versions of the game and told that the true payoffs were equally likely and would never be revealed publicly, but that the dictator could reveal them by clicking a button on the same screen. Both roles were informed that the dictator's decision of whether to reveal would be kept private. Thus, the dictator could remain ignorant of the payoffs, and the recipient would not know her information state. Failing to click the reveal button preserved the uncertainty, so 'not reveal' was the status quo or default choice.
3. *Active Choice*: This condition differed from the *Default NR* condition only in that the dictator was required to choose whether or not to reveal before the program proceeded to the outcome selection screen. There was no status quo or default choice.
4. *Default R*: This condition differed from the *Default NR* condition only in that the default information choice was to reveal the recipients' payoffs.
5. *Strategy Method*: This condition differed from the *Default NR* condition only in that the dictator entered her outcome choice as for each of the two payoff schemes, with the outcome determined by her choice in the game version actually being played. As in the *Default NR* condition, the dictator could reveal the payoffs by clicking a button on the

same screen.

The precise elicitation method and interface was not shown to the recipient in any condition, so the recipient was ignorant to the differences between conditions 2-5. Thus, all differences in behavior across these conditions are due to direct effects of the elicitation-method manipulation. The sessions lasted roughly 30 minutes. Each subject participated in only one condition and was not aware of the other conditions.

A total of 338 subjects participated across the five conditions, with exactly half (169) playing the role of dictator (Player X). On average participants earned $10.61, with dictators earning slightly more ($11.43) than recipients ($9.79). Sessions lasted approximately 20 minutes.

## 3 Results

Table 1 displays the results of the *Baseline* and *Default NR* conditions. While only 9 out of 36 (35%) dictators in the *Baseline* condition chose $(6, 1)$, in the *Default NR* condition 23 of 39 (59%) chose in a manner inconsistent with a preference for the fair outcome, either choosing to remain ignorant of the recipient's payoffs or, conditional on revealing and being in the *CI* game, choosing (6,1). The 24 percentage point difference is significant at the 3% level ($Z = 1.92$).[8] Thus, the main result of DWK is replicated.

Furthermore, the replication lines up closely with the original, differing by no more than ten percentage points along several key measures, with none of the differences being statistically significant. These measures include the baseline giving rate (.35 vs. .26, $Z = 0.59$), behavior inconsistent with a preference for fair outcomes (.59 vs. .53, $Z = 0.49$), the rate at which dictators remain ignorant of the recipient's payoffs (.54 vs. .44, $Z = 0.85$), the rate at which uninformed dictators choose $A : (6, 1)$ (.76 vs. .86, $Z = 0.69$), and the rate at of giving $(5, 5)$ conditional on knowingly being in the conflicting-interests game (.82 vs. .75, $Z = 0.36$). Thus, all the basic findings of the DWK moral wiggle-room game replicate.

[8] Unless otherwise noted, all hypothesis tests are one-tailed, pooled-sample tests of the difference of proportions.

| | | Replication | DWK |
|---|---|---|---|
| Baseline | $N$ | 26 | 19 |
| | Selfish ($A$) | .35 | .26 |
| Default NR | $N$ | 39 | 32 |
| | Inconsistent w/ fairness prefs. | .59 | .53 |
| | Ignorance rate | .54 | .44 |
| | $A$ given uninformed ($N$) | .76 (21) | .86 (14) |
| | Reveal & $CI$ giving ($N$) | .82 (11) | .75 (8) |

Table 1: The Baseline and Default NR conditions replicate the findings of DWK

Table 2 shows how the rate at which the dictator chose to avoid learning the recipient's payoffs changes with the default. Chosen ignorance decreases significantly from the *Default NR* to *Active Choice*, to *Default R* condition. The *Active Choice* condition features an ignorance rate of .25, which is significantly lower ($Z = 2.63$, $p < .01$) than the *Default NR* condition rate (.54). Switching to *Default R* pushes the ignorance rate even lower, to a level (.03) that is significantly different both from the *Default NR* condition ($Z = 4.39$, $p < .001$) and the *Active Choice* condition ($Z = 2.41$, $p < .01$).

| | Ignorance rate | $N$ |
|---|---|---|
| Default NR | .46 | 39 |
| Active Choice | .75 | 40 |
| Default R | .97 | 29 |

Table 2: Reversing the default choice significantly decreases the rate at which dictators choose to remain ignorant of the recipient's payoffs.

Table 3 displays results from the *Strategy Method* condition. The overall ignorance rate was .26, which is significantly greater ($Z = 2.46$, $p < .01$) than the .54 rate in the *Default*

*NR* condition. Among the 17 dictators who chose $A$ unconditionally (i.e. in both versions of the game), 29% chose ignorance. Among the 15 who chose $B$ only in the *CI* game, this rate varied little, with 27% choosing ignorance. These conditional choosers would learn the payoff state simply by observing their own payoff at the end of the session. The high rate at which they overcome the default choice in order to select 'reveal' suggests that activation cost or cost of overcoming the default is small relative to the desire to resolve uncertainty about the payoffs of the game *ex post*.

| | Ignorance rate | $N$ |
|---|---|---|
| Strategy Method | .26 | 35 |
| AA choosers | .29 | 17 |
| BA choosers | .27 | 15 |

Table 3: Participants in the *Strategy Method* condition avoid information at low rates, regardless of whether chosen strategy would eventually reveal the payoff state.

Table 4 summarizes the extent to which behavior was consistent with a preference for the the fair outcome $(5, 5)$. In the *Baseline* condition, choosing $(6, 1)$ is not consistent with a preference for the fair outcome. In the *Default NR*, *Active Choice*, and *Default R* conditions, a dictator who remains ignorant of the recipient's payoffs or knowingly chooses $(6, 1)$ in the *CI* game exhibits behavior inconsistent with a preference for the fair outcome. In the *Strategy Method* condition, I consider choosing the option that results in the recipient receiving only \$1 in either version of the game to be inconsistent with a preference for fair the fair outcome.

As described above, behavior in the *Default NR* condition is not consistent with the *Baseline* giving rate. However, neither the *Active Choice* nor the *Default R* condition features behavior that is less consistent with a preference for the fair outcome than in the *Baseline*. The inconsistency rate in the *Active Choice* condition is .35, exactly matching the *Baseline* rate, and the *Default R* features *less* inconsistent behavior (.28) than the *Baseline*.

The comparison with the *Strategy Method* condition is less clear. Out of 35 participants,

| | Inconsistent with fairness prefs. | $N$ |
|---|---|---|
| Baseline | .35 | 26 |
| Default NR | .59 | 39 |
| Active Choice | .35 | 40 |
| Default R | .28 | 29 |
| Strategy Method | .54 | 35 |

Table 4: Only in the *Default NR* condition is behavior inconsistent with the *Baseline* giving rate.

19 (54%) chose to give the recipient the lower payoff in at least one payoff state. This behavior is borderline significantly higher ($Z = 1.52$, $p < .07$) than the *Baseline* rate.[9] Thus, the *Default NR* condition is the only condition in which the amount of behavior inconsistent with a preference for the fair outcome is conclusively and significantly greater than in the *Baseline*.

## 4 Discussion and Conclusion

The sharp decline in the rate at which dictators choose to remain ignorant of the recipients' payoffs across the *Default NR*, *Active Choice*, and *Default R* conditions shows that dictators' information acquisition decisions are significantly affected by the default information choice. It is tempting to dismiss this result, and perhaps even the significance of the main DWK findings, as "merely" driven by status quo bias, a well-known and prevalent phenomenon, both in the lab (e.g. Samuelson and Zeckhauser (1988)) and in the field (e.g. Madrian and Shea (2001)). The default effect might be driven by a high physical or mental cost of selecting an additional choice, one that—unlike the outcome choice—can be avoided though passivity. The default choice may be interpreted as a suggestion on the part of the experimenter or as providing permission to act in a particular way.

[9] Excluding the one participant who chose the dominated option $(5, 1)$ in the aligned-interests game, the significance is even more tenuous ($Z = 1.31$, $p = .10$).

However, the *Strategy Method* condition shares the the same default as the *Default NR* condition, but almost three quarters of its participants overcome the default and actively choose to learn the recipients' payoffs. Because participants indicate contingent outcome choices on the same screen, revealing the payoffs serves only to resolve uncertainty *ex post*. Thus, while dictators' active pursuit of information is consistent with a general desire to resolve uncertainty or with a desire to learn the outcome of one's choice, it is markedly inconsistent with a simple status-quo-bias explanation for their avoidance of the same information *ex ante*, when it may have additional instrumental value. Instead, dictators appear to choose ignorance in order to avoid having to face an informed decision, but only when they can do so passively.

What is it about facing an informed decision that dictators would want to avoid? While there is little to suggest that dictators would want to avoid the $AI$ version decision, two different motivations might provide a reason to want to avoid the conflicting-interests decision. Each could explain why participants are willing to choose ignorance passively, but are reluctant to do so actively.

First, because their choice determines another person's payoff, dictators may feel obligated to choose the fair outcome, despite having a direct preference for the selfish outcome. Many studies have shown that a sense of responsibility for the welfare of others has an impact on behavior.[10] Dictators may be able to alleviate that responsibility by remaining ignorant, much in the way they can by delegating the decision to an intermediary[11] or by opting out of the allocation decision[12]. Indeed, Krupka and Weber (2010) find that the social consensus among experimental subjects who are given descriptions of the DWK baseline and 'moral wiggle room' game is that remaining ignorant is more socially-appropriate than choosing selfishly in either the baseline or while informed of being in the $CI$ game.

However, while unhelpful behavior on the part of the uninformed may be excused, a 'sin

[10] See for example, **?** or Charness and Jackson (2009).

[11] See for example Coffman (2009), Bartling and Fischbacher (2008), or Hamman, Loewenstien, and Weber (2009).

[12] See for example Lazear, Malmendier, and Weber (2009), Dana, Cain, and Dawes (2006), or Broberg, Ellingsen, and Johannesson (2007).

of commission' may be viewed more harshly than a 'sin of omission', and actively pursuing such ignorance may not be considered an abrogation of responsibility. This remains to be seen, though, and further research should seek to establish whether or not the active pursuit of ignorance does indeed spoil its responsibility-alleviating properties and is viewed as more inappropriate socially.

A second, related motivation is that image-concerned dictators are judged differently when ignorant than when informed. DWK is one of a handful of studies that largely inspired a shift in focus within the social-preference literature to models that incorporate concern for image.[13] Because self-sacrifice in the pursuit collective goals is socially valued, image-conscious decision-makers frequently face a conflict between material self-interest and the desire to maintain positive image. The tradeoff between self-interest and *social-image* can explain many patterns of giving behavior observed in laboratory and field experiments (Ariely, Bracha, and Meier 2009, Andreoni and Bernheim 2009, Grossman 2010) and changes in the information environment of potential recipients may alter this tradeoff enough to drive significant changes in behavior (Dana, Cain, and Dawes 2006, Grossman 2009b). However, because recipients in the DWK experiment did not observe the dictator's reveal decision, concern for social-image cannot explain why a dictator would avoid learning the recipient's payoffs.

A leading interpretation (espoused, for example by Benabou and Tirole (2006), Mazar, Amir, and Ariely (2008), Hamman, Loewenstien, and Weber (2009)) instead views the behavior observed by DWK as reflecting self-image concern and moral hypocrisy (Batson, Thompson, Seuferling, Whitney, and Strongman 1999): appearing moral to oneself without being so. By not revealing the payoffs, dictators can disregard the fact that they are not helping the recipient while simultaneously acting selfishly. Challenging this interpretation is the fact that Bayesian-rational self-signaling models such as Bodner and Prelec (2003), Benabou and Tirole (2006), and Grossman (2009b) cannot accommodate such self-deceptive

[13] See for example Benabou and Tirole (2006), Ariely, Bracha, and Meier (2009), Andreoni and Bernheim (2009), Grossman (2009b).

behavior. It remains a mystery how participants can fool themselves into thinking that they are not blameworthy if they do not learn the state. Furthermore, the direct experimental tests derived from such models have found little support for the hypothesis that self-signaling is a major driver of giving behavior (Grossman 2009b, Grossman 2009a).

In the experiments described in this paper, overcoming the default draws attention to the dictator's choice. This added attention may limit the dictator's ability to self-deceive, destroying the value of strategically ignoring the recipient's payoffs, which otherwise could be used to avoid facing the decision featuring conflicting dictator and recipient interests. Thus, while opportunities to create and exploit moral wiggle-room limit fair-minded behavior, environmental or psychological variables that draw attention to one's own behavior reinforce the motivation that leads people to choose fair outcomes. Further planned experiments will investigate whether language and stimuli that draw attention to the dictator's choice—without changing the default—are similarly successful at neutralizing strategic ignorance and reinforcing fair-minded behavior.

Related studies have found similar patterns of experiment participants avoiding dictator decisions, even at a cost (Lazear, Malmendier, and Weber 2009, Dana, Cain, and Dawes 2006, Broberg, Ellingsen, and Johannesson 2007). Krupka and Weber (2010) find such opt-out is viewed as more appropriate socially than entering the dictator game and playing selfishly. Further research might investigate whether it is also the case that environmental factors drawing attention to the participant's choice mitigate the extent to which participants avoid allocation decisions.

## References


Andreoni, J., and B. D. Bernheim (2009): "Social Image and the 50-50 Norm," forthcoming.

Ariely, D., A. Bracha, and S. Meier (2009): "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially," *American Economic Review*, Forthcoming.

Bartling, B., and U. Fischbacher (2008): "Shifting the Blame: On Delegation and Responsibility," Discussion paper.

Batson, C., E. Thompson, G. Seuferling, H. Whitney, and J. Strongman (1999): "Moral hypocrisy: appearing moral to oneself without being so," 77(3), 525–537.

Benabou, R., and J. Tirole (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96(5), 1652–1678.

——— (2010): "Identity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, forthcoming.

Bodner, R., and D. Prelec (2003): "Self-signaling and Diagnostic Utility in Everyday Decision Making," in *The Psychology of Economic Decisions, vol. Rationality and Well-being*, ed. by I. Brocas, and J. Carillo, chap. 6. Oxford University Press.

Bolton, G. E., and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Broberg, T., T. Ellingsen, and M. Johannesson (2007): "Is Generosity Involuntary," *Economics Letters*, 94(1), 32–37.

Camerer, C. (2003): *Behavioral Game Theory*. Princeton University Press.

Charness, G., and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, 74(6), 1579–1601.

Charness, G., and M. Jackson (2009): "The Role of Responsibility in Strategic Risk-Taking," *Journal of Economic Behavior and Organization*, 69, 241 – 247.

Charness, G., and M. Rabin (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3), 817–869.

Coffman, L. (2009): "Intermediation Reduces Punishment (and Reward)," working paper.

Dana, J., D. M. Cain, and R. Dawes (2006): "What you don't know won't hurt me: Costly (but quiet) exit in a dictator game," *Organizational Behavior and Human Decision Processes*, 100(2), 193–201.

Dana, J., R. A. Weber, and J. X. Kuang (2007): "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," *Economic Theory*, 33, 67–80.

Dufwenberg, M., and G. Kirchsteiger (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47, 268–298.

Falk, A., and U. Fischbacher (2006): "A Theory of Reciprocity," *Games and Economic Behavior*, 54(2), 293–315.

Fehr, E., and K. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10(2), 171–178.

Greiner, B. (2003): "An Online Recruitment System for Economic Experiments," *Forschung und wissenschaftliches Rechnen*, 63, 79–93.

Grossman, Z. (2009a): "A Paper-based Self-signaling Test," working paper.

——— (2009b): "Self-signaling versus social-signaling in giving," UC Santa Barbara Department of Economics Working Paper Series.

——— (2010): "Social-signaling Under Anonymity: Beliefs-based Altruism or Rule-rationality?," working paper.

Hamman, J. R., G. Loewenstien, and R. A. Weber (2009): "Self-interest through delegation: An additional rationale for the principal-agent relationship," *American Economic Review*, forthcoming.

Krupka, E. L., and R. A. Weber (2010): "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?," IZA Discussion Papers 3860, Institute for the Study of Labor (IZA).

Lazear, E. P., U. Malmendier, and R. A. Weber (2009): "Sorting and Social Preferences," working paper.

Madrian, B., and D. F. Shea (2001): "The Power of Suggestion: Interia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116(4), 1149 – 1187.

Mazar, N., O. Amir, and D. Ariely (2008): "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance," *Journal of Marketing Research*, 45(6), 633–644.

Rabin, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281–1302.

Samuelson, W., and R. Zeckhauser (1988): "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1(1), 7–59.

## Experiment

INSTRUCTIONS: PLEASE READ VERY CAREFULLY. IF YOU HAVE A QUESTION, PLEASE RAISE YOUR HAND AND WAIT FOR ASSISTANCE.

This is an experiment in the economics of decision-making. Several research institutions have provided funds for this research. You will be paid for your participation in the experiment. The exact amount you will be paid will depend on your and/or others' decisions. Your payment will consist of the amount you accumulate plus a $5 participation bonus. You will be paid privately in cash at the conclusion of the experiment.

If you have a question during the experiment, raise your hand and an experimenter will assist you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Please put away all outside materials (such as book bags, notebooks, cellphones) before starting the experiment. Participants violating the rules will be asked to leave the experiment and will not be paid.

OK

## Experiment

Description of the Game

In this experiment, each of you will play a game with one other person in the room. Before playing, we will randomly match people into pairs. The grouping will be anonymous, meaning that no one will ever know which person in the room they played with. Each of you will be randomly assigned a role in this game. Your role will be player X or player Y. This role will also be kept anonymous. The difference between these roles will be described below. Thus, exactly one half of you will be a Player X and one half a Player Y. Also, each of you will be in a pair that includes exactly one of each of these types.

The game your pair will play will be like the one pictured below. Player X will privately choose one of two options: "A" or "B". Player Y will not make any choice. Both players will receive payments based on the choice of Player X. The numbers in the table are the payments players receive. The payments in this table were chosen only to demonstrate how the game works. In the actual game, the payments will be different. For example, if player X chooses "B", then we should look in the bottom square for the earnings. Here, Player X receives 3 dollars and Player Y receives 4 dollars.

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 1 | 2 |
| B | 3 | 4 |

OK

**Experiment**

Check your understanding

At this point, to make sure that everyone understands the game, please answer the two questions below.

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 1 | 2 |
| B | 3 | 4 |

In this example, if Player X chooses "B" then:

Player X receives $ [ ]

Player Y receives $ [ ]

In this example, if Player X chooses "A" then:

Player X receives $ [ ]

Player Y receives $ [ ]

OK

baseline

Experiment

You are Player X. To make your choice, please select one of the options below, then click OK to confirm your choice.

I choose
- A
- B

OK

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | 1 |
| B | 5 | 5 |

## Experiment

Two possible versions of the actual game

The actual game you will play will be one of the two pictured below. Notice that both games are the same except that two of Player Y's payments have been switched. Note that in both games, Player X gets his or her highest payment of $6 by choosing A. In the game on the left, this gives Player Y his or her lowest payment of $1. In the game on the right this gives Player Y his or her highest payment of $5. In both games, if Player X chooses B, he or she gets a lower payment of $5. In the game on the left, this gives Player Y the highest payment of $5. In the game on the right, this gives Player Y the lowest payment of $1.

Note that you will not know which of the games that you are playing, but for Player X, the payments will be identical. The only thing that differs are the payments for Player Y.

Which of these games will you actually play? That was determined randomly by the computer at the beginning of the experiment, with each game being equally likely. However, we will not reveal publicly which game you are actually playing. Player X can choose to find out which game is being played if he or she wants to do so by clicking a button. This choice will be anonymous, thus Player Y will not know if X knows which game is being played. However, Player X is not required to find out and may choose not to. Player Y will not have this option. When the game ends, we will pay each player privately.

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | 1 |
| B | 5 | 5 |

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | 5 |
| B | 5 | 1 |

OK

**Experiment**

Check your understanding

To make sure that everyone understands the game, please answer the two questions below. Remember that each of the two possible versions of the game shown below are equally likely.

Which option gives Player X his or her highest payment in both games?
- A
- B

If Player X chooses B, then Player Y receives
- $5
- $1
- either $5 or $1

OK

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | 1 |
| B | 5 | 5 |

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | 5 |
| B | 5 | 1 |

# Thinking delay

Time remaining 56

Thinking time

Please take a minute to think about your decision. After one minute you will be told whether you are Player X or Player Y. If you are Player X, you will then make your decision.

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | 1 |
| B | 5 | 5 |

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | 5 |
| B | 5 | 1 |

Default NR

Experiment

If there are no further questions, we will begin the game.

You are Player X. To make your choice, click the corresponding button below. If you wish to reveal which of the two games is actually being played, click "Reveal Game".

Player Y will not find out whether or not you learned the actual version of the game.

I choose:
- A
- B
- Reveal Game

OK

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | ? |
| B | 5 | ? |

# Active Choice

If there are no further questions, we will begin the game.

You are Player X. You will choose one of the options below. But first, please decide whether you wish to reveal which game is actually being played.

Player Y will not find out whether or not you learned the actual version of the game.

Reveal game? ○ No ○ Yes

OK

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | ? |
| B | 5 | ? |

# Active Choice

Experiment

You are Player X. To make your choice, please select one of the options below, then click OK to confirm your choice.

Player Y will not find out whether or not you learned the actual version of the game.

I choose ○ A ○ B

OK

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | ? |
| B | 5 | ? |

Default R

Experiment

If there are no further questions, we will begin the game.

You are Player X. Click the OK button to proceed to the next screen and make your choice. If you do not wish to reveal which of the two games is actually being played, first click the "Do Not Reveal Game" button.

Player Y will not find out whether or not you learned the actual version of the game.

- (•) Proceed
- ( ) Do Not Reveal Game

OK

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | ? |
| B | 5 | ? |

# Strategy Method

Experiment

If there are no further questions, we will begin the game.

You are Player X. Though you have not been told the actual version of the game that is being played, please maker your choice for each version by clicking the corresponding button below. If the actual game is the one on the left, your payment and the payment of Player Y will be determined by the choice you indicated for the left game. If the actual game is the one on the right, the payments will be determined by the choice you indicated for the game on the right.

For the game on the left, I choose ○ A ○ B

For the game on the right, I choose ○ A ○ B

To confirm your choices, please click OK when you are done. If you would to see which of the two games you actually played, click the "Reveal Game" button before clicking OK.

○ Reveal Game

OK

| Player X chooses | Player X receives | Player Y receives |
|---|---|---|
| A | 6 | ? |
| B | 5 | ? |