Department of Economics

HENLEY
UNIVERSITY OF READING

Focal Points, Gender Norms and Reciprocation in Public Good Games

by
David Zetland and Marina Della Giusta

2011
089

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

FOCAL POINTS, GENDER NORMS AND RECIPROCATION IN PUBLIC GOODS GAMES

DAVID ZETLAND AND MARINA DELLA GIUSTA

Abstract. We examine the impact of information regarding other people's choices on individual choice in a public good experiment with two separate treatments. In the `implicit` treatment, subjects do not see the average contribution of others in their group, but they can calculate it from the information available. In the `explicit` treatment, subjects see the average contribution of others in their group. If subjects are rational calculating agents as suggested in mainstream economic theory there should be no difference in observed behavior across treatments: agents should use all available information to make decisions.

What we see instead is quite different and consistent with the presence of social norms: first, players change their behavior in response to the change in displayed information; second, changes in individual behavior produce identical group outcomes, in terms of total payoffs or efficiency across the two treatments. How does this happen? The display of the average contribution of others results in behavior consistent with a focal point (Schelling, 1960), i.e., more subjects behave as reciprocators (conditioning their contributions on the contributions of others), and fewer behave as cooperators or free-riders (unconditionally contributing a lot or a little, respectively). This change in behavior differs by gender: women behave similarly to men when they see the average contribution by others; when they cannot, they behave differently, favoring unconditional strategies of free-riding or cooperation. Men's behavior, in contrast to women's adaption, does not adjust to social cues, as suggested by Croson and Gneezy (2009).

1. Introduction

Over the last few years the experimental literature has provided a wealth of empirical evidence in support of the idea that we are fundamentally social beings and our behavior is influenced by that of others (Fehr et al., 2009). Social cues can be a powerful source of motivation: gene-culture co-evolution models suggest that groups tend to prefer members to share values in order to work more productively together and punish transgression of social norms (Boyd and Richerson, 1989; Bowles, 1998, 2001). The evidence from game theory

Date: March 31, 2011. 7,600 words.
Key words and phrases. Public goods, focal points, social norms, gender, experiments. JEL: D0.
Address. Zetland `dzetland@gmail.com` is at Wageningen University; Della Giusta (`m.dellagiusta@reading.ac.uk`) is at University of Reading. Thanks to Josh Abbott, Antoine Champetier de Ribes, Catherine Eckel, Dan Houser, Stephan Kroll, Rob Kurzban, John List, Travis Lybbert, Harold Margolis and participants at the 2007 and 2008 ESA for their comments. We thank the A.G. Giannini Foundation and the Department of Agricultural and Resource Economics at the University of California, Davis for financial support.

indicates that agents like to punish transgressors, and that social pressure can improve reciprocation from selfish agents (for a review of the evidence see Fehr et al. (2009)). Behavioral experiments in both psychology and economics have also provided support to the presence of framing effects: presentation changes the way people respond to a problem (Kahneman and Tversky, 2000). Research on framing in psychological games (games in which payoffs depend on beliefs and not just actions) by Dufwenberg et al. (2010) shows that social influences and framing interact with each other through belief formation: frames influence beliefs about others' beliefs (rather than their actions) and that might be sufficient to change player's own chosen actions.

Social norms have long featured in explanations of individual and group behavior by economists, from the early work on social norms and conformism by Akerlof (1980) and Jones (1984) to the recent contributions by Akerlof and Kranton (2000, 2005); Corneo and Jeanne (2009, 2010). Akerlof's theory of social custom states that social norms are endogenously determined and depend on the proportion of people who believe in them. As individuals differ both in their beliefs and the willingness to conform to other people's beliefs, the resulting range of social codes is consistent with multiple equilibria.

The degree of conformity with social norms is individual, but systematic variation by gender has been observed in the burgeoning literature on the effect of gender on preference formation and mutation. For example Kamas et al. (2008) find that there are gender differences in altruistic behavior with women giving more than men in anonymous giving to charity and increasing total contributions in mixed-sex pairings. A recent review of the evidence from experimental economics by Croson and Gneezy (2009) suggests that although women are not necessarily more altruistic than men, they are more likely to be affected by social clues on appropriate behavior.In Della Giusta et al. (2011a), we found that conformity with social norms is higher for women than men in the UK, and that whilst part of the difference in life satisfaction between women and men in the UK can be attributed to the different activities men and women carry out, a substantial part can be explained by the fact that women and men give different weights to satisfaction with different life dimensions, effectively constructing systematically different utility functions (Della Giusta et al., 2011b).

We examine the impact of information regarding other people's choices on individual choice in a public good experiment with two separate treatments. In the `implicit` treatment, subjects do not see the average contribution of others in their group, but they can calculate it from the information available. In the `explicit` treatment, subjects see the average contribution of others in their group. If subjects are rational calculating agents as suggested in mainstream economic theory there should be no difference in observed behavior across treatments: agents should use all available information to make decisions.

What we see instead is quite astonishing. First, players change their behavior in response to the change in displayed information (more exactly, one group of players behaves differently in one environment than a different group of similar players in a different environment). Second, the mix of changes in individual behavior produces identical group outcomes in terms of total payoffs or efficiency. In an everyday context, consider the example of a man who changes his movements with the terrain, to maintain a steady pace. Now consider that the man in our public goods game is actually a women — women's actions change with the information display in our two different treatments, to maintain the same outcome.

How does this happen? The display of the average contribution of others results in behavior *consistent* with a focal point (Schelling, 1960), i.e., more subjects behave as reciprocators (conditioning their contributions on the contributions of others), and fewer behave as cooperators or free-riders (unconditionally contributing a lot or a little, respectively).[1] Women are responsible for the change in the mix of types. They behave similarly to men in `Explicit`, but more women pursue unconditional strategies of free-riding or cooperation in `Implicit`. The display of information matters, and it matters because female behavior differs with the information environment.

[1]Unfortunately, the experiment does not create a *pure* focal point (i.e., one equilibrium in many equilibria on which subjects converge with the knowledge that others will also converge) because the display of the average contribution of others facilitates an existing strategy of reciprocation. Nevertheless, we use "focal point" in the loose sense, to describe a feature that subjects converge upon.

The next section describes the experimental treatments. Section 3 presents results, Section 4 has the discussion, and Section 5 concludes. The appendices have game instructions, a discussion of typing methods, and an analysis of subject-pool heterogeneity.

2. Treatments

In each run of a public goods game (PGG), each player in a group of n splits his endowment (e) between the group's public account and his private account. A subject's total earnings are the sum of his private earnings plus a fraction of the public account. The researcher sets the value of this fraction (the Marginal Contribution Ratio or MCR) somewhere between $\frac{1}{n}$ and 1.00 — the extremes leaving participants indifferent to, respectively, contribution or non-contribution to the public account.

Given that all players benefit from contributions to the public account, the rational strategy is to free-ride by contributing nothing to the public account and e to the private account — leading to a Nash equilibrium in which all subjects receive e (Ledyard, 1995). The social-welfare maximizing strategy is for all subjects to contribute e to the public account and nothing to their private accounts, so that each subject receives $MCR(n * e) \geq e$. We define "group efficiency" as average subject earnings (from public and private accounts) divided by maximum possible earnings. A group of free-riders would have an efficiency of $e/MCR(n*e)$ that will range from $\frac{100}{n}$ percent (when MCR is 1.00 and the public good creates the greatest welfare gains) to 100 percent (when MCR is $\frac{1}{n}$ and the public good produces no welfare gains).

By observing each player's contributions to the public account, we can also classify him as a "cooperator" (contributing a lot to the public account, independently of what others do), "free-rider" (contributing little or nothing, independently of what others do), or "reciprocator" (contributing more when others do).

2.1. **Game Procedure.** This PGG is modeled on the sequential-contribution design of Kurzban and Houser (2005) — hereafter KH — where subjects start with an initial, simultaneous contribution and then get one or more chances to update it before the game ends at a random point. In `Implicit` (as in KH), the average contribution of others is not shown. In `Explicit`, the average is shown. Besides this difference [**in BOLD below**], `Implicit` and `Explicit` are identical to each other (but not to KH; see Footnote 11). Appendix A has a copy of the instructions. The games had the following steps:

(1) Subjects were randomly placed in groups of four or five at the beginning of each run.[2]

(2) In period zero, all subjects had 20 seconds to make a simultaneous, initial contribution to the public account from their 50-token endowments.[3] Participants understood that their remaining tokens were allocated to their private accounts.

Subjects knew they had at least one opportunity to confirm/change their period zero contribution. Provisional contributions were final only when the run ended. Period zero contributions were non-binding cheap talk, but decisions after period zero were payoff relevant because the run could end at any point.

(3) After period zero, contributions were sequential, i.e., each player had ten seconds to change or confirm his contribution while the rest of the group waited. In `Implicit`, each subject saw his contribution, the total, and the number of people in his group and "knew" the average contribution of others to be (total–own contribution)/(groupsize–1); see Figure 1. **In `Explicit`, he sees the average contribution of others; see Figure 2.**

[2]We used groups of four and five to maximize "yield" among student recruits. There were two `Explicit` sessions where students were in groups of four or five. If we omit those sessions from the results, the shares of types among the six remaining `Explicit` sessions are not statistically different from shares in all eight sessions.

[3]We use the following terminology: In each session, subjects played five runs of the PGG; each run had 7–32 rounds (individual decisions), grouped into two to eight periods after period zero. After period zero, each period had the same number of rounds as the maximum number of subjects in any group, i.e., a group of four would have four rounds (decisions) in one period. (In mixed groups of four and five, subjects in groups of four waited while the five-player groups finished. Because subjects were already waiting for others in their group, additional waiting for the five-player groups did not affect the flow of the game.) The run ended — even in the middle of a period — when the maximum number of rounds (between 7 and 32) was reached.

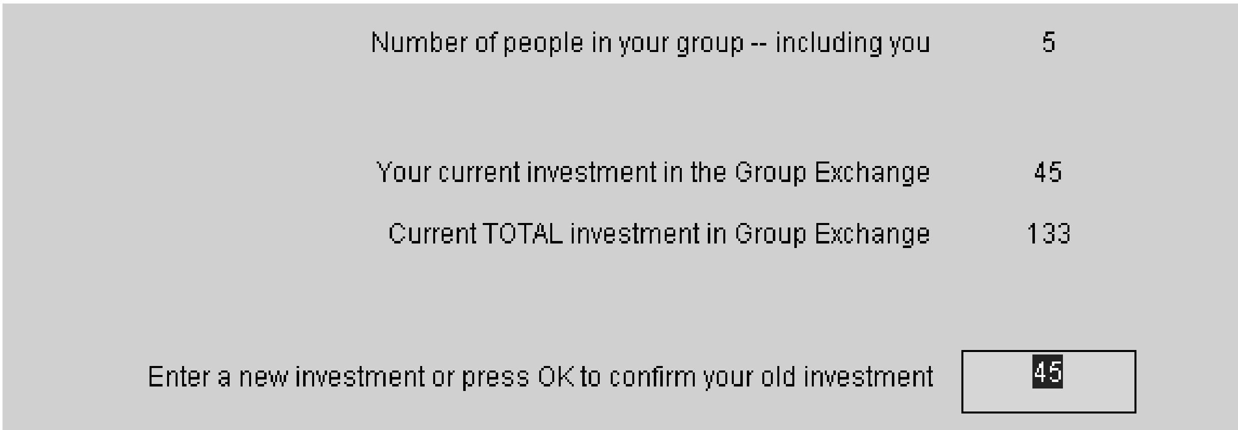

Figure 1. Screenshot from `Implicit`.

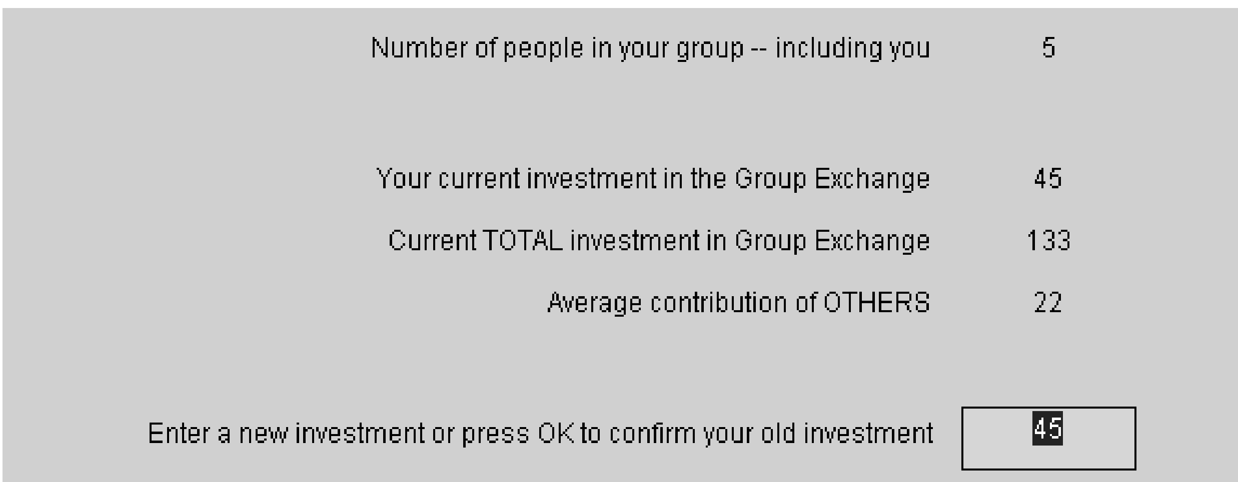

Figure 2. Screenshot from `Explicit`, which adds "Average contribution of OTHERS."

(4) Each group's public account total was updated, the next round began, and the next member of the group could change/confirm his contribution. Rounds and periods ran without signal or interruption.

(5) Updating continued for an unknown, random number of rounds until the run ended, contributions were finalized, and subjects saw their payoffs from public and private contributions.[4] Each subject received one token for each token in his private account and 0.5 token ($MCR = 0.5$) for each token in the group account.

[4]Average contributions were biased upwards by the limited number of rounds and $1/n$ probability of any given run ending after a player's contribution decision. This bias applies to both treatments, so we ignore it.

(6) In each session, there were five runs of the game, each ending after a quasi-random number of rounds.[5] Participants were randomly shuffled into groups and randomly ordered within groups at the beginning of each run and played in the same order for that run. They knew they were in new groups, but they did not know the number of rounds in each run or number of runs in the session.

2.2. **Typing Subjects.** Each subject played the game for five runs and made one to eight payoff-relevant decisions per run (period zero contributions are ignored) with some players making more decisions than others.[6] Thus, each player in each session has 26–28 "average contribution of others/own contribution" datapoints.[7] Players are typed using these datapoints and the following equation:

$$x_{igt} = \alpha_i + \beta_i \bar{x}_{igt} + \epsilon_{igt}, \tag{1}$$

where x_{igt} is the contribution of person i in group (run) g in round t, $\bar{x}_{igt}$ is the average contribution of other group members observed by i; α_i and β_i are individual-specific parameters to be estimated, and ϵ_{igt} is a mean-zero disturbance term ($\sim N(0, \sigma_i^2)$) that controls for group effects (g) and trend effects (t).

Each individual's type depends on α_i and β_i values estimated through an individual OLS regression of Equation (1).[8] Given point estimates of $\hat{\alpha}$ and $\hat{\beta}$, KH's classification rules for type are as follows:

[5]The first run took 16 rounds, the second was seven rounds, then 23, 32 and 32 rounds. These counts (from KH) were used in all sessions.

[6]When the number of rounds divided by the number of players is not an integer the number of observations is not equal; e.g., four subjects playing seven rounds would mean that three subjects played two rounds each and one played a single round.

[7]Although it is possible to confound player behavior with behavior of the group, this problem is minimized by reshuffling the players into different groups five times during the session.

[8]Since KH's classification method uses point estimates for individuals' parameter values (i.e., ignoring statistical significance; see Appendix B), the error term is ignored. This method therefore ignores both group effects and trend effects. The impact of the former is minimized by shuffling players between groups. Although we ignore trend effects, there is the possibility that learning matters across the five runs. When we retype players using data from runs 2–5 (i.e., controlling for "learning"), 21/296 (seven percent) of players change from one type to another, and aggregate shares of types shift by 0–3 percent. The between-treatment difference of reciprocator shares actually increases. Since this result agrees with (and reinforces) our main result, we also ignore trend.

Cooperator: $\hat{\beta} \geq 0$ and $\hat{\alpha} \geq 25$, i.e., a cooperator's estimated contribution is non-decreasing in the average contribution of others and always at least 25 (of 50) tokens.[9]

Free-rider: $\hat{\beta} \geq 0$ and $\hat{\alpha} + \hat{\beta}(50) < 25$, i.e., a free-rider's estimated contribution is non-decreasing in the average contribution of others but stays below 25 tokens.

Reciprocator: $\hat{\beta} \geq 0$, $\hat{\alpha} < 25$, and $\hat{\alpha} + \hat{\beta}(50) \geq 25$, i.e., a reciprocator's estimated contribution is non-decreasing in the average contribution of others, below 25 tokens when the average contribution of others is zero, and at least 25 tokens when the average contribution of others is 50.

No Type: $\hat{\beta}_i < 0$. Players who give less when others give more are classified as "no type" and ignored in further analysis.

Figure 3 shows datapoints for subjects who were classified as cooperators and free-riders; Figure 4 shows subjects classified as reciprocators. Each panel shows all data for one player in one session; the fitted line matches regression output, i.e., $\hat{x}_i = \hat{\alpha}_i + \hat{\beta}_i \hat{\bar{x}}_i$. Note that each dot is a $(\bar{x}_{igt}, x_{igt})$ pair that records the average contributions of others to the public account (independent variable on x-axis) and how much that subject puts in the public account (dependent variable on y-axis).

KH's method of classifying subjects with linear OLS point estimates is fast and easy to use (one can "type" players with a few observations), but some worry that it ignores potentially important factors. First, point estimates ignore the error structure of ϵ_{igt}; second, type may not fit a linear profile; and third, Tobit is more appropriate for typing censored observations. Appendix B shows that these effects do not have a significant impact on our typing results. With Occam's Razor in hand, we use KH's method instead of other, more complicated means of classification discussed in, e.g., El-Gamal and Grether (1995); Houser et al. (2004); Houser and Winter (2004).

[9]KH used 25/50 tokens (50 percent) as a cut-off between "free-riders" who never give more than half their endowment (contributing less than 25 even when others average 50) and cooperators who always give more than half (contributing more than 25 even when others average zero). In ultimatum games (one player decides how to split an endowment and his partner decides to reject — leaving both with nothing — or accept the split), the modal offer is fifty/fifty; see Camerer and Thaler (1995).

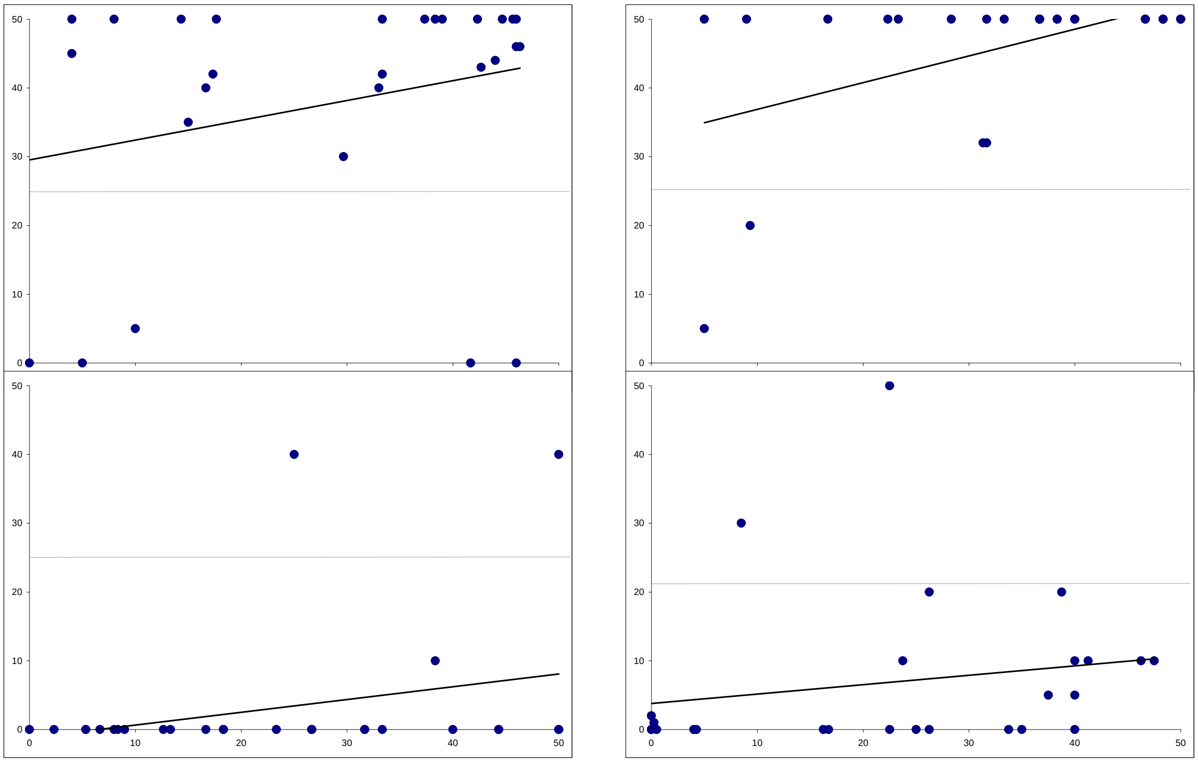

FIGURE 3. Subjects typed as cooperators (top) and free-riders (bottom). Contributions (y-axis) are in response to the (implicit or explicit) average contribution of others (x-axis).

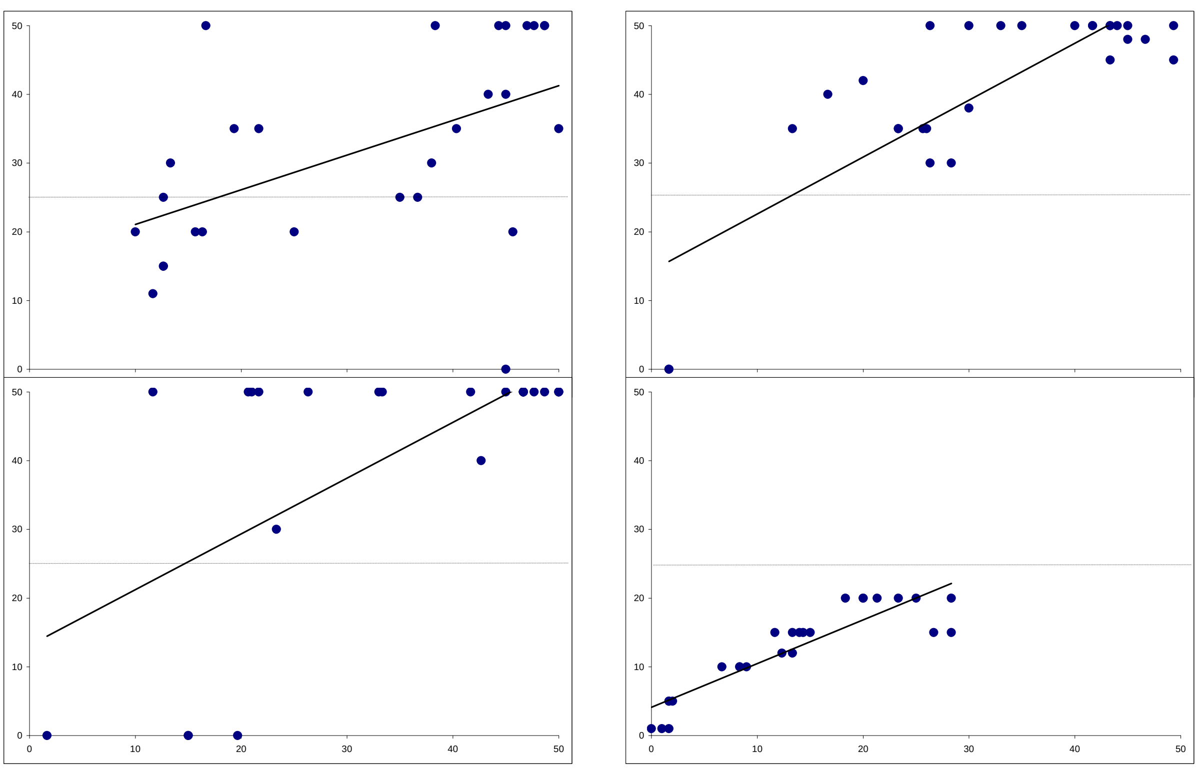

FIGURE 4. Subjects typed as reciprocators

TABLE 1. Descriptive statistics of different sessions.

	Kurzban-Houser	`Implicit`	`Explicit`
Location	U. of Arizona	UC Davis	UC Davis
Dates	2002	April 2007	October 2006
Sessions	4	8	8
Students	3x24, 1x12	20x6, 16x2	20x4, 16x2, 19x1, 13x1
Total Students	84	152	144
Students Classified	81	139	136
Share Male	N/A	42%	57%
Share Economists	N/A	37%	52%

2.3. **Hypotheses.** KH determine each subject's type by comparing his contribution to the public account to the average contribution of others in his group — a number he never sees. In other words, KH assume subjects *know* the average contribution of others. If that is true, subject behavior (and thus type) should be the same in `Implicit` and `Explicit`. The following hypotheses test that idea:

$\mathbf{H}_0^c$**:** The shares of cooperators in `Implicit` and `Explicit` are the same.

$\mathbf{H}_0^f$**:** The shares of free-riders in `Implicit` and `Explicit` are the same.

$\mathbf{H}_0^r$**:** The shares of reciprocators in `Implicit` and `Explicit` are the same.

2.4. **Session Details.** Each treatment in the between-subject design was repeated in eight sessions at a UC Davis computer lab.[10] Those sessions differ in many ways from KH (different software, instructions, group sizes and time limits). Tables 1 and 2 display results from our treatments with KH's results, but they are not comparable.[11]

Unclassified students are classified as "no type" and dropped from analysis. The impacts of different shares of males and economists are discussed in Appendix C.

Each session began after subjects signed legal consent/disclosure forms, received their anonymous participant number, and heard directions in Appendix A. Participants played

[10]Although the between-subject design has problems with subject fixed effects (addressed in Appendix C), a within-subject design is not feasible because subjects (who could easily identify the difference between treatments) could bring experience from one treatment to the next, invalidating any results from a sequential, within-subject treatment design.

[11]KH use Visual Basic, and we use z-Tree (Fischbacher, 2007). `Explicit` has groups of four or five students (e.g, the 19 student session has three groups of five students and two groups of four students), but `Implicit` and KH have groups of four. Decision times are limited in `Implicit` and `Explicit` but not in KH.

TABLE 2. MLOGIT-adjusted shares of each type by treatment

	KH		Implicit		Explicit	
	% Share	95% CI	% Share	95% CI	% Share	95% CI
Cooperators	13.6	12.8–14.4	10.8	10.4–11.2	4.4	4.1–4.7
Free-riders	21.0	20.0–22.0	26.5	25.9–27.1	11.8	11.3–12.2
Reciprocators	65.4	64.3–66.6	62.7	62.0–63.4	83.8	83.3–84.4

either `Implicit` or `Explicit` for five runs. After each run, they were reshuffled into new groups of four or five.[12] After completing the PGG and another game (not reported here), subjects answered a questionnaire.[13] Finally, each player received an anonymous cash payment in proportion to his performance. The average payment was about $15. Total session length was less than 1.5 hours.

3. Results

Table 2 — adjusted for subject-pool heterogeneity; see Appendix C — shows that `Implicit` had fewer reciprocators, more cooperators and more free-riders than `Explicit`. Wilcoxon Mann-Whitney rank-sum tests comparing the shares of types (e.g., shares of reciprocators in eight `Implicit` sessions to shares in eight `Explicit` sessions) reject equal shares at the one percent level for all three types, i.e.,

$\mathbf{H}_0^c$**:** The shares of cooperators in `Implicit` and `Explicit` are the same. **Reject**.

$\mathbf{H}_0^f$**:** The shares of free-riders in `Implicit` and `Explicit` are the same. **Reject**.

$\mathbf{H}_0^r$**:** The shares of reciprocators in `Implicit` and `Explicit` are the same. **Reject**.

[12]Such reshuffling means that players in `Explicit` (which had two sessions with five-player groups) would have been in a mix of four and five-player groups. Given the same 0.5 MCR, the greater incentive to cooperate should have resulted in more cooperators and reciprocators, but the share of cooperators in these two sessions was zero; the shares of reciprocators and free-riders were 88 and 12 percent, respectively compared to 83 and 11 percent in the six `Explicit` sessions with four-player groups and 84 and 12 percent for all `Explicit` sessions. (Players in five-player groups who saw larger total contributions — relative to four-player groups — also saw average contributions of others in their group, which would weaken the impression that others were being more generous than they really were.)

[13]After `Explicit`, subjects participated in an auction game; after `Implicit`, they participated in a different public goods game that tested behavior under varying incentives to cooperate, i.e., tournament, in-group/out-group, etc. While these games may have had different impacts on questionnaire answers, we cannot isolate these effects.

4. Discussion

> ... manipulations in the lab experiment can yield drastically different measures of individual propensities. This result does not necessarily imply that preferences are labile. Rather, we view such data as evidence that when critical elements of the situation change, behavior will change in predictable ways.
>
> — Levitt and List (2007, p. 164)

Our results offer significant implications for experimental design and policy implementation. Firstly, if subject behavior depends not just on information but on the display of information, then experimenters should be cautious about the design, execution and interpretation of their treatments. Secondly, our evidence suggests that the link between underlying preferences and revealed preferences is not as tight as rational action theory would like it to be. For example, subjects with unconditional preferences as cooperators or free-riders should also behave unconditionally. But framing that leads subjects to behave as reciprocators may lead us to mistakenly assume that we have elicited underlying preferences when we have actually induced behavior (and type) as a response to circumstances (Margolis, 2007, Chapter 9).[14] Smith (2003) calls this "ecological rationality," i.e., the notion that we will do the right thing at the right time, and the focal point effect discussed here is consistent with that rationality.[15]

In the large experimental literature on framing, social interaction and focal points, the following articles touch on the present discussion: Dietrich et al. (2001) show that boundedly-rational players pay attention to displayed information — not what they *should* know. In Berg et al. (2005), subjects change from risk averse to risk seeking as framing (the same

[14]Tversky and Kahneman (1986, p. s273) defines framing as "the language of presentation, on the context of choice and on the nature of the display."

[15]Bowles (1998) makes the more radical argument that preferences are endogenous, while McCubbins and Weller (2009) find that player behavior is rational when it's conditioned on their belief about the actions of others.

gamble) changes. Croson (2007) finds that subjects try to match a displayed median contribution more than they try to match the minimum or maximum contribution.[16] Cooper and Rege (2008) find that players' actions tend to polarize around the same choices — a result they attribute to social interaction and peer effects. Czap and Ovchinnikova (2008) arbitrarily make one player a leader in a public goods game and display the leader's contribution; followers' contributions are positively affected by the leader but not vice-versa. In studies that challenge the robustness of effects consistent with a focal point, Crawford et al. (2008) find that a focal point effect deteriorates when payoffs are asymmetric, and Brosig et al. (2008) find that players who behave as unselfish "types" in a dictator game do not do so when the game is repeated weeks later.[17]

The explicit display of the average contribution of others "causes" more subjects to behave as reciprocators and fewer to behave as cooperators or free-riders. From a policy perspective, this result implies that the display of information on others' actions combined with conformism to social norms can guide agent behavior. Our finding that overall efficiency does not change can be seen either as evidence that such a program of display will not make us worse off, as a group. Alternatively, it may indicate that such a program will not increase social welfare. That latter argument ignores the possibility that a "nudging" policy (that's easy to ignore) may increase overall utility from the process of arriving at a final result. It's not hard to see that participation in a group with a larger share of reciprocators may result in greater procedural utility than participation in a group with free-riders and cooperators (Tyler, 1990; Benz, 2004; Frey, 2005). Many people dislike free-riders; cooperators attract a combination of disdain and envy (Fehr and Gächter, 2000; Rodriguez-Sickert et al., 2008; Herrmann et al., 2008).

[16]She did not include mean contribution, but subject contributions have a similar correlation with observed median and unobserved mean.

[17]In the game discussed here, players see the average contribution of others, which reproduces the peer-interaction structure of Croson (2007) and Cooper and Rege (2008) — but not the leader structure of Czap and Ovchinnikova (2008) — and a change in results seen in Dietrich et al. (2001) and Berg et al. (2005). Because the game uses symmetric payments and repeats only once, it does not test the results of Crawford et al. (2008) or Brosig et al. (2008), but these treatments are not designed to defend the robustness of focal points or types.

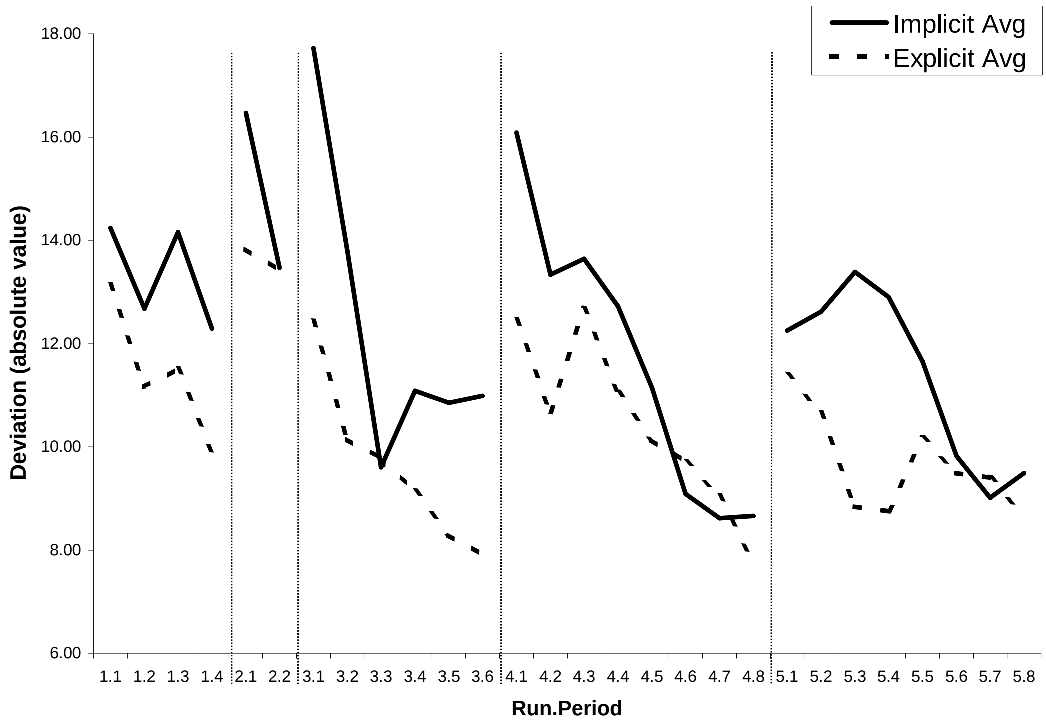

FIGURE 5. Reciprocators' contributions are closer to the mean in `Explicit`. Vertical lines separate the five runs — each lasting two to eight periods.

Besides this type result, we should see — a là the "match the average finding in Croson (2007) — other evidence that subjects are responding to a focal point, e.g., individual contributions should converge on the average contribution of others in `Explicit`. Figure 5 shows the absolute value of the difference between the average contribution of subjects typed as reciprocators and the average contribution of others in those subjects' groups.[18] A t-test of the data behind the figure (i.e., comparing average deviations by `run.period` between `Implicit` and `Explicit`) rejects the null hypothesis that differences are equal at the one-percent level. Individual contributions in `explicit` are closer to the group average.

For the remainder of this discussion, we will explore the factors driving these results — starting with the role of gender (Section 4.1) before trying to understand how the gender effect may result from lazy and/or strategic behavior (Sections 4.2 and 4.3).

[18]These data are from 14 sessions with groups of four; two `Explicit` sessions with groups of five are omitted.

TABLE 3. Percentage shares of types, by gender and treatment

	`Implicit`		`Explicit`	
	Females	Males	Females	Males
Cooperators	12	9	5	4
Free-riders	38	17	10	13
Reciprocators	49	74	85	83

Before proceeding, let us clarify how and why we got here. Our original impetus for running `Implicit` was to understand why `Explicit` results varied from KH's (implicit) results. Since we could not be sure that we were controlling for all relevant differences between KH and `Explicit`, we ran `Implicit` six months later. Although we were able to test (and reject) the null hypotheses, our desire to understand *why* the results differed (and control for subject-pool heterogeneity) led us to run the multinomial logit described in Appendix C. Those results made it clear that gender (female behavior) was driving results. And now let us proceed.

4.1. **Gender and Behavior.** Table 3 shows that a majority (51 percent) of females in `Implicit` act as (unconditional) cooperators or free-riders while a majority (74 percent) of males choose a conditional (reciprocator) strategy. This discrete descriptive result is reinforced by the continuous density functions for male and female $\hat{\alpha}$ and $\hat{\beta}$ values (i.e., intercept and slope, respectively from Equation 1). Two-sample Kolmogorov-Smirnov tests of all subjects *fail* to reject the hypothesis that the distribution of $\hat{\alpha}$ values from `Implicit` and `Explicit` are the same for males and females. The same test also fails to reject statistical equivalence for male $\hat{\beta}$ values but not for female $\hat{\beta}$ values. Put differently, females have a different response to others (p-value < 0.001) in each treatment.

Figure 6 makes this difference clear. The left panel displays kernel-density functions for female $\hat{\beta}$ values, which are skewed to the right in `Implicit` (thick line) but not in `Explicit` (thin line). For males (right panel), $\hat{\beta}$ values are similar in both treatments. Thus we see how different female behavior explains the differences in results between the two treatments. (This result is robust to multi-variate controls; see Appendix C.) So why do women behave differently from men?

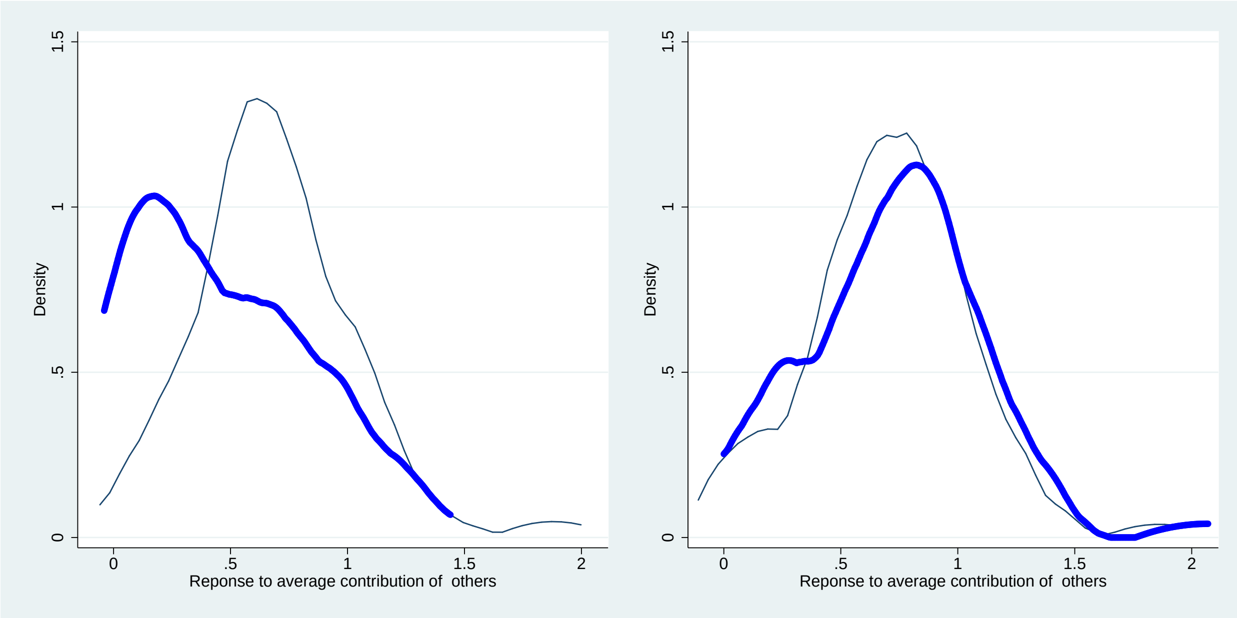

FIGURE 6. $\hat{\beta}$ values for females (left panel) and males (right panel) for `Implicit` (thick line) and `Explicit` (thin line).

First, let us dismiss the explanation that subjects facing the ten-second clock in `Implicit` made errors in calculating their contribution — errors that resulted in them being misclassified as free-riders or cooperators when they were "really" reciprocators. Data on individual decisions show that few were made at the last second. Instead, it seems that subjects who wanted to reciprocate had adequate time to make their calculation, and many males (74 percent) did behave as reciprocators. Females, on the other hand, were more likely to pursue the unconditional strategies of cooperation or free-riding.

Second, we have to consider the so-called "Larry Summers Critique," i.e., the idea that women are not as good at math as men. Although Benjamin et al. (2006) find that men and women calculate equally well, it is possible that women do not *bother* to make the calculations necessary for reciprocation in `Implicit` — an explanation that echoes the comparative advantage finding that girls do as well as boys in math and better in reading — leading them to pursue non-mathematical activities (Guiso et al., 2008). Thus, females may reciprocate less often because they prefer to do otherwise, either because it is easier or because they have other plans. Let us consider these "lazy" and "strategic" explanations.

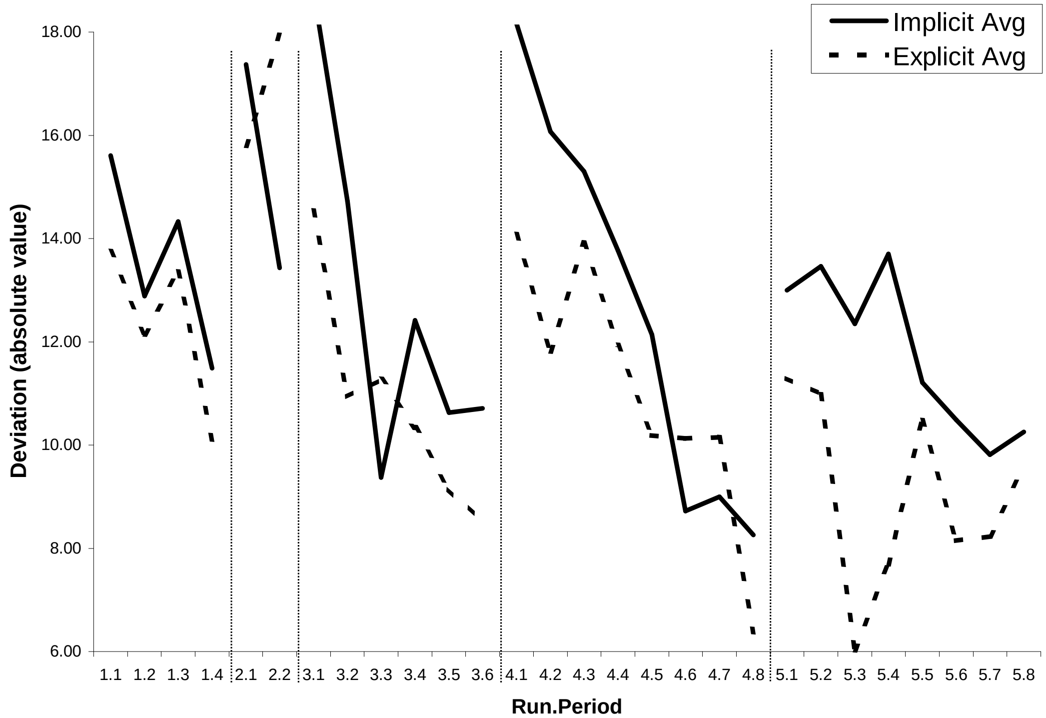

FIGURE 7. Average absolute value of difference between a male subject's contribution and the average contribution of others in his group

TABLE 4. Average reciprocator deviation from contribution of others, by gender and treatment, in first and last period of each run.

	`Implicit`		`Explicit`	
	Females	Males	Females	Males
First Period	14.03	16.59	11.79	13.89
Last Period	11.14	10.83	8.07	10.50
Decrease	2.89	5.76	3.72	3.39
% Decrease	21%	35%	32%	24%

4.2. **Lazy Reciprocators.** Although both male and female reciprocators contribute closer to the average in `Explicit` (Figures 7 and 8 display Figure 5 results by gender), their behavior patterns differ in each treatment. Table 4 shows that male reciprocators "work harder" to match the average contribution of others in `Implicit` while women work harder than men (and finish closer to the average contribution of others) in `Explicit`.[19]

[19]There is little evidence of learning differences between genders. In `Explicit`, men and women have a similar deviation in period 1 of run 1 (13.75 and 13.58, respectively) and in period 1 of run 5 (11.29 and 11.45, respectively). In `Implicit`, men (women) start at 15.60 (12.78) and drop to 13.00 (11.45).

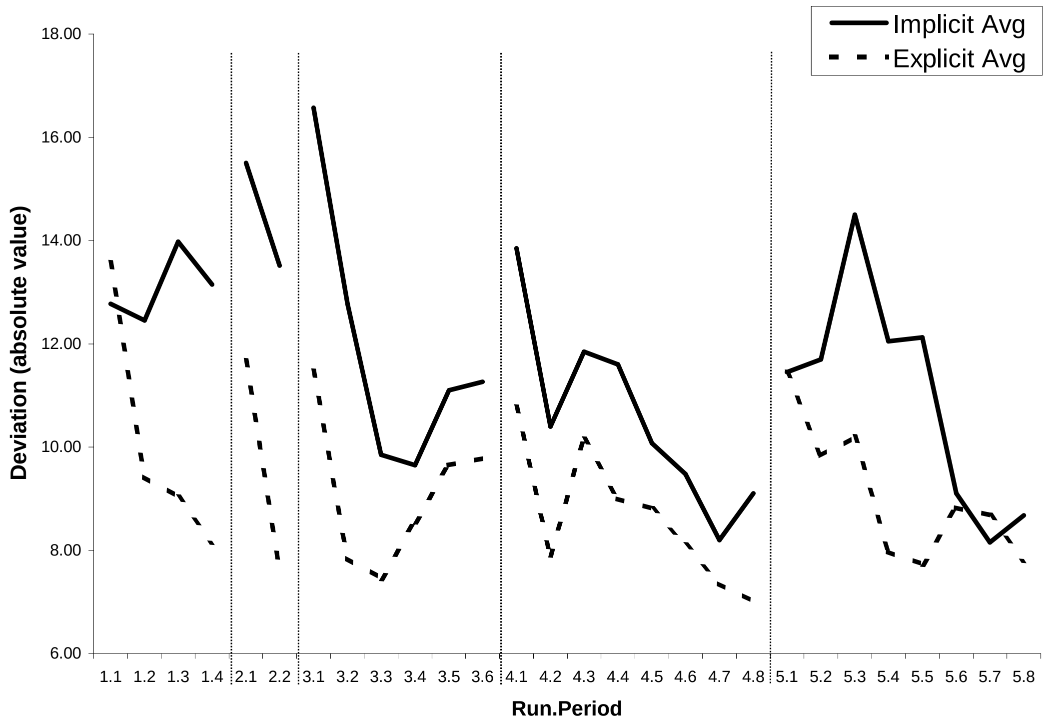

FIGURE 8. Average absolute value of difference between a female subject's contribution and the average contribution of others in her group.

These results echo those of Cadsby and Maynes (1998a,b), who find that groups of females are significantly better at coordinating around a common equilibrium than groups of males, and Niederle and Yestrumskas (2008), who find that women equal men in competition that involves performance-related feedback, i.e., the case in `Explicit`.

Because `Implicit` lacks an equilibrium, women without a focal point may take the easier way out (i.e., adopting an unconditional strategy). Males, on the other hand, may work relatively harder because they are more competitive (Niederle and Vesterlund, 2007; Croson and Gneezy, 2009) or want to beat the game (Charness and Levin, 2008).[20]

4.3. **Strategic Reciprocators.** The smaller deviation from the mean in `Explicit` might also result from players' strategic use of a focal point to coordinate strategies. If players are indeed pursuing interdependent strategies, the implication is that they — as a group — will reach some form of steady-state efficiency that cannot be displaced by another group. Put

[20]Gneezy et al. (2009) find that female competitiveness is driven by culture (nurture), not gender (nature).

differently, the behavior of interacting subjects should produce a mix of types (cooperators, free-riders and reciprocators) that results in efficient outcomes appropriate to the treatment. The data support this idea: Despite a dramatic change in the mix of types, average player efficiencies — at 68.1 percent in `Implicit` and 67.8 percent in `Explicit` — are not statistically different (p-value = 0.64).[21] This result — group earnings in each treatment are statistically identical — echoes Smith's "ecological rationality," i.e., that players change their behavior to fit the environment.

The driver of this result appears to be women; they behave differently in different treatments. Such a change has been found elsewhere and attributed to women being more strategic or sensitive to changes in the information environment (Nowell and Tinkler, 1994; Seguino and Lutz, 1996; Benjamin et al., 2007; Leon-Mejia and Miller, 2007; Croson and Gneezy, 2009). Since `Explicit` has a focal point, it is more likely that women will pursue a reciprocation strategy when others are likely to do the same.[22]

5. Conclusion

Public goods experiments have failed to find consistent differences in the behavior of men and women — perhaps because of experimental heterogeneity (Eckel and Grossman, 2008; Croson and Gneezy, 2009). These treatments clarify gender differences by altering a single experimental element (the implicit or explicit display of the average contribution of others) to show how gender and design interact to produce gender-specific behavior in `Implicit` but gender-neutral behavior in `Explicit`. In `Explicit`, the shares of female and male reciprocators are, respectively, 85 and 83 percent. In Implicit, these shares drop to 49 and 74 percent, respectively (shares of both free-riders and cooperators rise). The difference is significant for females but not for males.

[21] Efficiency does not vary by gender either.

[22] Do they change "type" in a strategic way to equalize payouts? Although we have no evidence of that (not least because treatments are between-subject), it's possible that these results reflect emergent social coordination.

The best explanation for the difference in results is that women change their behavior when the informational environment changes responding to a visible cue (focal point) that informs them about the social norm. Interestingly, the change in types does not result in different average payouts, indicating that aggregate behavior in each treatment is optimal in the sense of being evolutionarily stable. This suggests that women's behavior balances men's behavior in a way that maintains stability for the whole population. More research is needed to establish whether this result applies to wider pool of subjects, but the policy implications can be quite important in devising incentive systems based on compliance with social norms.

These results also contradict a common (perhaps naive) assumption that preferences reliably map to behavior — underscoring the importance of understanding the impact of design on experimental results.

References

Akerlof, G. and Kranton, R. (2005). Social divisions within schools: how school policies can affect students' identities and educational choices. In Barrett, C., editor, *The Social Economics of Poverty*, pages 188–213. Routledge, New York.

Akerlof, G. A. (1980). A Theory of Social Custom – of which Unemployment may be One Consequence. *Quarterly Journal of Economics*, 94(4):749–775.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *Quarterly Journal of Economics*, 115(3):715–53.

Benjamin, D. J., Brown, S. A., and Shapiro, J. M. (2006). Who Is "Behavioral?" Cognitive Ability and Anomalous Preferences. SSRN Working Paper 675264.

Benjamin, D. J., Choi, J. J., and Strickland, A. J. (2007). Social Identity and Preferences. NBER Working Paper 13309.

Benz, M. (2004). Introducing Procedural Utility: Not only What, but also How Matters. *Journal of Institutional and Theoretical Economics*, 160(3):377–401.

Berg, J., Dickhaut, J., and McCabe, K. (2005). Risk Preference Instability across Institutions: A Dilemma. *Proceedings of the National Academy of Sciences*, 102(11):4209–4214.

Bowles, S. (1998). Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions. *Journal of Economic Literature*, 36(1):75–111.

Bowles, S. (2001). *Social Dynamics*, chapter Individual interaction, group conflicts and the evolution of preferences. Brookings Institution Press, Washington, D.C.

Boyd, R. and Richerson, P. J. (1989). The evolution of indirect reciprocity. *Social Networks*, 11:213–236.

Brosig, J., Riechmann, T., and Weimann, J. (2008). Selfish in the End? An Investigation of Consistency and Stability of Individual Behavior. *Working Paper*. `http://mpra.ub.uni-muenchen.de/2035/`.

Cadsby, C. B. and Maynes, E. (1998a). Choosing Between a Socially Efficient and Free-Riding Equilibrium: Nurses versus Economics and Business Students. *Journal of Economic Behavior & Organization*, 37(2):183–192.

Cadsby, C. B. and Maynes, E. (1998b). Gender and Free Riding in a Threshold Public Goods Game: Experimental Evidence. *Journal of Economic Behavior & Organization*, 34(4):603–620.

Camerer, C. and Thaler, R. H. (1995). Anomalies: Ultimatums, Dictators and Manners. *Journal of Economic Perspectives*, 9(2):209–219.

Charness, G. and Levin, D. (2008). The Origin of the Winner's Curse: A Laboratory Study. *AEJ: Microeconomics*, 1(1).

Cooper, D. J. and Rege, M. (2008). Social Interaction Effects and Choice Under Uncertainty: An Experimental Study. *Working Paper*. `http://tinyurl.com/5dfe9n`.

Corneo, G. and Jeanne, O. (2009). A theory of tolerance. *Journal of Public Economics*, 93(5–6):691–702.

Corneo, G. and Jeanne, O. (2010). Symbolic values, occupational choice, and economic development. *European Economic Review*, 54(2):237–251.

Crawford, V. P., Gneezy, U., and Rottenstreich, Y. (2008). The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures. *American Economic Review*, 98(4):1443–58.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.

Croson, R. T. A. (2007). Theories of Commitment, Altruism and Reciprocity: Evidence from Linear Public Goods Games. *Economic Inquiry*, 45(2):199–216.

Czap, H. J. and Ovchinnikova, N. V. (2008). Are Leaders Willing to Make Sacrifices for an Environmental Public Good? Are Followers More Than Just Followers? *Working Paper*.

Della Giusta, M., Jewell, S., and Hashimzade, N. (2011a). Why Care? Social Norms and the Supply of Unpaid Care. *University of Reading Working Paper*.

Della Giusta, M., Jewell, S., and Kambhampati, U. (2011b). His and hers: exploring gender puzzles and the meaning of life satisfaction. *Feminist Economics*, forthcoming.

Dietrich, J. R., Kachelmeier, S. J., Kleinmuntz, D. N., and Linsmeier, T. J. (2001). Market Efficiency, Bounded Rationality, and Supplemental Business Reporting Disclosures. *Journal of Accounting Research*, 39(2):243–268.

Dufwenberg, M., Gaetcher, S., and Hennig-Schimdt, H. (2010). The Framing of Games and the Psychology of Play. *University of Nottingham CEDEX Discussion Paper*, 2010-16.

Eckel, C. and Grossman, P. (2008). Differences in the Economic Decisions of Men and Women: Experimental Evidence. In Plott, C., editor, *Handbook of Experimental Results*. Elsevier, New York.

El-Gamal, M. and Grether, D. M. (1995). Are People Bayesian? Uncovering Behavioral Strategies. *Journal of the American Statistical Association*, 90(432):1137–1145.

Fehr, E., Brown, M., and Zehnder, C. (2009). On reputation: a microfoundation of contract enforcement and wage rigidity. *Economic Journal*, 119:333–353.

Fehr, E. and Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4):980–994.

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

Fischbacher, U., Gachter, S., and Fehr, E. (2001). Are People Conditionally Cooperative? Evidence from a Public Goods Experiment. *Economics Letters*, 71(3):397–404.

Frey, Bruno S.and Stutzer, A. (2005). Beyond Outcomes: Measuring Procedural Utility. *Oxford Economic Papers*, 57:90–111.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica*, 77(5):1637–1664.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, Gender, and Math. *Science*, 320:1164–1165.

Gunnthorsdottir, A., McCabe, K., and Smith, V. (2002). Using the Machiavellianism Instrument to Predict Trustworthiness in a Bargaining Game. *Journal of Economic Psychology*, 23(1):49–66.

Herrmann, B., Thoni, C., and Gachter, S. (2008). Antisocial Punishment Across Societies. *Science*, 319(5868):1362–1367.

Houser, D., Keane, M., and McCabe, K. (2004). Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm. *Econometrica*, 72(3):781–822.

Houser, D. and Winter, J. (2004). How Do Behavioral Assumptions Affect Structural Inference? Evidence from a Laboratory Experiment. *Journal of Business & Economic Statistics*, 22(1):64–79.

Jones, S. R. (1984). *The Economics of Conformism*. Basil Blackwell, New York.

Kahneman, D. and Tversky, A. (2000). *Choices, values and frames*. Cambridge University Press, New York.

Kamas, L., Preston, A., and Baum, S. (2008). Altruism in individual and joint-giving decisions: what's gender got to do with it? *Feminist Economics*, 14(3):23–50.

Kurzban, R. and Houser, D. (2005). An Experimental Investigation of Cooperative Types in Human Groups: A Complement to Evolutionary Theory and Simulations. *Proceedings of the National Academy of Sciences*, 102(5):1803–1807.

Ledyard, J. O. (1995). Public Goods: A Survey of Experimental Research. In Kagel, J. H. and Roth, A. E., editors, *Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ.

Leon-Mejia, A. and Miller, L. M. (2007). The Devil Is in the Details: Sex Differences in Simple Bargaining Games. Jena Economic Research Paper 2007-069.

Levitt, S. D. and List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *Journal of Economic Perspectives*, 21(2):153–174.

Margolis, H. (2007). *Cognition and Extended Rational Choice*. Routledge, New York.

McCubbins, M. D. and Weller, N. (2009). Irrational Beliefs, Rational Actions: Understanding Subjects' Behavior in a Laboratory Experiment. *SSRn Working Paper*, 1443980.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M. and Yestrumskas, A. H. (2008). Gender Differences in Seeking Challenges: The Role of Institutions. NBER Working Paper 13922.

Nowell, C. and Tinkler, S. (1994). The Influence of Gender on the Provision of a Public Good. *Journal of Economic Behavior & Organization*, 25:25–36.

Rodriguez-Sickert, C., Guzman, R., and Cardenas, J. C. (2008). Institutions influence preferences: Evidence from a common pool resource experiment. *Journal of Economic Behavior & Organization*, 67(1):215–227. Forthcoming JEBO.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.

Seguino, Stephanie, S. T. and Lutz, M. A. (1996). Gender and Cooperative Behavior: Economic Man Rides Alone. *Feminist Economics*, 2(1):1–21.

Smith, V. L. (2003). Constructivist and Ecological Rationality in Economics. *American Economic Review*, 93(3):465–508.

Tabellini, G. (2008). The Scope of Cooperation: Norms and Incentives. *Quarterly Journal of Economics*, 123(3):905–950.

Tversky, A. and Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *Journal of Business*, 59(4):S251–S278.

Tyler, T. R. (1990). *Why People Obey the Law*. Princeton University Press, Princeton, NJ.

Appendix A. PGG Instructions

This is a game of group and individual investment behavior.

- You are in a group of 4 with 3 others, chosen at random. (If you are in a GROUP OF FIVE, you will find out during the game.)
- You have an endowment of 50 tokens to invest. Others have the same endowment.
- You invest your tokens in the Individual Exchange and the Group Exchange.
- Your earnings depend on how you and your group invest tokens.
- 50 tokens = $0.75

Every token you invest in the **Individual Exchange** returns one token in earnings **to you** only.

Every token you invest in the **Group Exchange** returns 0.5 tokens in earnings **to every member of your group, including yourself**. *It does not matter* who invests in the Group Exchange — everyone gets a return from every token invested in the Group Exchange, whether or not they invested.

Your task is to maximize your earnings by choosing how many of your tokens to invest in the Group Exchange. (Remaining tokens go to the Individual Exchange.) Examples:

	1	2	3
Your Group Exchange investment	0	50	30
Your Individual Exchange investment	**50**	**0**	**20**
If others' Group Exchange investments total	90	110	0
. . . total Group Exchange investment is . . .	$0 + 90 = 90$	$50 + 110 = 160$	$30 + 0 = 30$
. . . and everyone's Group Exchange return is	$90/2 =$ **45**	$160/2 =$ **80**	$30/2 =$ **15**
Your total earnings (in tokens) are	$50 + 45 = 95$	$0 + 80 = 80$	$20 + 15 = 35$

Game Timing.

(1) All members of your group start with a simultaneous investment in the Group Exchange (**Round 1**). Click "Continue" after you enter your choice. *You only have 20 seconds to click.* A countdown clock is in the top-right corner of your screen.

(2) In **Round 2 and thereafter**, you will (one person at a time) see the number of people in your group (either 4 or 5), the TOTAL investment in the Group Exchange, [**Explicit:** "and the average investment of others in your group"]. You will change or confirm your Group Exchange investment and click "Continue." *You only have 10 seconds to click.* If you take too long, your choice does not change.

(3) The opportunity to see the total and change/confirm passes from person to person in your group for an unknown, random number of rounds until the run ends, and all investments are final. You will have *at least* one opportunity to change/confirm your investment. Although you must wait while the decision passes around your group, try to pay attention so as to not to miss your turn.

(4) When each run ends, you will see your investment, the total investment in the Group Exchange, your earnings from the current run, and your cumulative earnings.

(5) When the game repeats, players are randomly reshuffled into new groups and the final round changes to a new, random number.

Appendix B. Typing Subjects

Table 5 compares shares of types resulting from KH's OLS method (described in Section 2.2) and other OLS methods (discussed below). To reduce confusion and maintain focus on typing *methods* (as opposed to outcomes), shares of each type under each method are shown for the `Explicit` treatment only.

Table 5. Subject types by estimation method

OLS Regression	Shares of Type (Percent) Coop.	Free-Rider	Recip.	Hump	No Type (count)
all coefficients	4	12	84		8
signif. coeff. only	3	27	69		1
quadratic (all coeff.)	4	10	72	14	6

B.1. **Statistical Significance.** OLS regressions give estimates of $\hat{\alpha}_i$ and $\hat{\beta}_i$, and these coefficients are statistically insignificant for some individuals. KH use all point estimates in their typing — regardless of statistical significance — and that method is used here. Using all estimates means ignoring error structure, but this cost to accuracy is more than compensated by avoiding an even greater problem — bias in typing.

Bias is introduced when insignificant estimates are set to zero because zero values of $\hat{\beta}$ are associated with free-riders. Put differently, assigning individuals with coefficients that are not significantly different from zero to the free-rider group would overstate the importance of free-riders. We can see this effect in Table 5, where the share of free-riders rises from 12 percent with point estimated coefficients to 27 percent with significant coefficients only.

B.2. **Quadratic Form.** KH's typing method forces each player's actions to fit a linear form, and it is not hard to imagine that some players may play a different strategy, e.g., increasing contributions up to a certain point and then decreasing them. Fischbacher et al. (2001) find that 14 percent of subjects have such a "hump-shaped" contribution profile.

Allowing for quadratic variation in Equation (1) gives us:

$$x_{igt} = \alpha_i + \beta_i \bar{x}_{igt} + \gamma_i (\bar{x}_{igt} - \bar{x}_i)^2 + \epsilon_{igt}, \tag{2}$$

where $\bar{x}_i$ is the average contribution of others for all rounds and γ_i is an additional parameter to be estimated. $\bar{x}_i$ is used in $(\bar{x}_{igt} - \bar{x}_i)^2$ to increase variation in the quadratic relationship and reduce problems with collinearity.

When typing with Equation 2, six subjects are excluded as "no type" (meaningful negative β and γ coefficients). Of the 138 remaining subjects, 14 percent (as in Fischbacher et al. (2001)) are humped types (positive β and negative γ coefficients); 4 percent are cooperators; 10 percent are free-riders; and 72 percent are reciprocators.

How do we interpret these results? The main shift (compared to the linear-only method) is from reciprocators to hump-shaped. In Table 5, we see that the share of reciprocators in OLS falls from 84 percent to 72 percent in OLS-quadratic, which has 14 percent hump-shaped types. This result implies that some subjects typed as reciprocators with a linear approximation are really contributing fewer tokens when the average contribution to the public account is above 28–30 tokens.

We do not use quadratic estimates because classification of types is more arbitrary (without statistical significance as a filter, most subjects have *some* value for γ, which introduces confusion as to who is a free-rider, reciprocator, etc.), and the only difference appears to be a transfer from reciprocators to hump-shaped types. (It would be more troubling if free-riders and cooperators also moved to hump-shaped types.)

B.3. **Tobit Model.** Finally, there is the much larger issue of contributions that are censored at upper and lower boundaries in the estimation model (46 percent of contribution decisions — x_{igt} values — are 0 or 50). Since an OLS estimate of the relationship between censored values of x_{igt} and $\bar{x}_{igt}$ will produce inconsistent estimates, a Tobit model would probably be more accurate. We do not use Tobit because we would need to create a typing scheme around different estimates, which is not the point of this paper.

Appendix C. Controlling for Subject Pools

It is possible that subject-pool heterogeneity affected the mix of types — recall the differing shares of males and economists in Table 1. A multinomial logit model is used to test the effect of gender, field of study, etc. on type. Note that type depends on coefficients from individual OLS regressions of own contribution on average contribution of others. MLOGIT estimations use type as the dependent variable and previously-unused subject characteristics as independent variables.

Table 6 has regression output, and Table 7 shows the estimated marginal effects of independent variables on the probability of being classified as a given type.[23]

Note that the **Trust Index** (TI) RHS variable is calculated from the answer to four yes/no questions: "People generally do the right thing," "I find it better to accept others for what they say and appear to be," "I am doubtful of others until we know they can be trusted," and "I almost always believe what people tell me."[24] Yes answers are added (+1,+1,-1,+1) to get individual TI values between -1 and 3. Although the correlation between answers to these questions and types in a PGG is not perfect (e.g., someone who trusts others may be greedy, i.e., a free-rider), there is good reason to believe that stated preferences should match revealed preferences. Put differently, TI values should be higher for cooperators, lower for free-riders and somewhere in-between for reciprocators; see, e.g., Tabellini (2008).

The estimated marginal effects indicate that a participant in `Explicit` is 6 percent (insignificant) less likely to be a cooperator, 25 percent less likely to be a free-rider, and 30 percent more likely to be a reciprocator. Field of study (economics, quantitative or non-quantitative) is not statistically-correlated with behavior or types. Males in `Implicit` are 3 percent (insignificant) less likely than females to be cooperators, 17 percent less likely to be free-riders, and 20 percent more likely to be reciprocators.[25] These results reflect the importance of gender and treatment on behavior, types and outcomes.

An adjustment to `Implicit` results that reflects different shares of males in subject pools (42 percent in `Implicit` and 57 percent in `Explicit`) would shift about 3 percent of all subjects (males are 20 percent more likely to reciprocate, but they were 15 percent fewer) from free riders (29.5 falls to 26.5%) to reciprocators (59.7 rises to 62.7%). Table 8 shows results before adjustment; Table 9 shows these adjusted shares, which we consider to be a more accurate reflection of the mix of types in our subject pool.

Importantly, it appears that males are not affected by the difference in treatments. When we rerun the MLOGIT with a dummy to compare male and female behavior between treatments (i.e., replace dummies for `Explicit`, `Male x Explicit` & `Male x Implicit` with dummies for `Male`, `Female x Explicit` & `Male x Explicit`), coefficients on `Male` are significant for free-riders and reciprocators and significant for all types for `Female x Explicit` but *not* significant for `Male x Explicit`. Thus it appears that treatment effects manifest entirely through females.

[23]A multinomial probit has a similar fit with slightly-higher statistical significance. The addition of a non-interacted male dummy does not change results.

[24]Pre-testing included 20 "Machiavellian" questions (see Gunnthorsdottir et al. (2002)), and the four with the lowest collinearity were retained.

[25]We use `Male x Implicit` and `Male x Explicit` instead of the more conventional `Male` and `Male x Explicit` to get clean estimates of between-treatment gender differences.

Table 6. Multinomial Logit of type on subject characteristics. Type 3 (Reciprocator) and non-quantitative major (anything but economics or quantitative majors such as engineering, sciences and MBA/Finance) omitted.

	coefficient	p-value
Type = 1 (Cooperator)		
Explicit Treatment	-1.53	0.14
Male x Implicit	-0.83	0.19
Male x Explicit	0.16	0.86
Econ x Implicit	0.43	0.61
Econ x Explicit	-0.30	0.78
Quant x Implicit	0.79	0.29
Quant x Explicit	0.55	0.61
Age	0.06	0.75
Household size	-0.32	0.11
Years in major	0.30	0.29
Experimental Experience	-0.66	0.34
Trust Index	0.33*	0.10
Intercept	-2.77	0.48
Type = 2 (Free-Rider)		
Explicit Treatment	-1.76***	0.01
Male x Implicit	-1.51***	0.00
Male x Explicit	0.17	0.77
Econ x Implicit	0.70	0.17
Econ x Explicit	0.16	0.81
Quant x Implicit	-0.02	0.97
Quant x Explicit	0.17	0.84
Age	0.08	0.60
Household size	0.04	0.61
Years in major	0.02	0.94
Experimental Experience	0.29	0.45
Trust Index	-0.20	0.14
Intercept	-2.21	0.44
N=275	Log-likelihood = -184.56	
Pseudo $R^2 = 0.119$	$\chi^2_{(20)} = 49.71$	

Significant at the 1% (***), 5% (**) and 10% (*) levels

Table 7. Marginal effects after Multinomial Logit

	dy/dx	p-value
	Type = Cooperator	
Explicit Treatment	-0.057	0.29
Male x Implicit	-0.026	0.28
Male x Explicit	0.006	0.89
Econ x Implicit	0.014	0.76
Econ x Explicit	-0.015	0.73
Quant x Implicit	0.051	0.39
Quant x Explicit	0.031	0.69
Age	0.002	0.81
Household size	-0.016*	0.07
Years in major	0.015	0.29
Experimental Experience	-0.031	0.21
Trust Index	0.019*	0.06
	Type = Free-Rider	
Explicit Treatment	-0.250**	0.02
Male x Implicit	-0.170***	0.00
Male x Explicit	0.024	0.79
Econ x Implicit	0.115	0.24
Econ x Explicit	0.028	0.79
Quant x Implicit	-0.013	0.87
Quant x Explicit	0.019	0.88
Age	0.012	0.62
Household size	0.008	0.42
Years in major	0.000	0.98
Experimental Experience	0.052	0.41
Trust Index	-0.034*	0.10
	Type = Reciprocator	
Explicit Treatment	0.307***	0.00
Male x Implicit	0.196***	0.00
Male x Explicit	-0.031	0.75
Econ x Implicit	-0.129	0.21
Econ x Explicit	-0.125	0.91
Quant x Implicit	-0.038	0.69
Quant x Explicit	-0.051	0.72
Age	-0.014	0.56
Household size	0.008	0.55
Years in major	-0.014	0.66
Experimental Experience	-0.022	0.74
Trust Index	0.015	0.49

dy/dx is for discrete change of dummy variable from 0 to 1

Significant at the 1% (***), 5% (**) and 10% (*) levels

TABLE 8. Shares of each type by treatment

	KH		Implicit		Explicit	
	% Share	95% CI	% Share	95% CI	% Share	95% CI
Cooperators	13.6	12.8–14.4	10.8	10.4–11.2	4.4	4.1–4.7
Free-riders	21.0	20.0–22.0	29.5	28.9–30.1	11.8	11.3–12.2
Reciprocators	65.4	64.3–66.6	59.7	59.0–60.4	83.8	83.3–84.4

TABLE 9. MLOGIT-adjusted shares of each type by treatment

	KH		Implicit		Explicit	
	% Share	95% CI	% Share	95% CI	% Share	95% CI
Cooperators	13.6	12.8–14.4	10.8	10.4–11.2	4.4	4.1–4.7
Free-riders	21.0	20.0–22.0	26.5	25.9–27.1	11.8	11.3–12.2
Reciprocators	65.4	64.3–66.6	62.7	62.0–63.4	83.8	83.3–84.4