Draft: Comments welcome

Other Regarding Preferences: A Selective Survey of Experimental Results*

David J. Cooper
Florida State University

John H. Kagel
Ohio State University

4/11/2009

*Research support from the National Science Foundation grant SES-0451981 is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Outline

Introduction

There has been an enormous amount of experimental research devoted to "other regarding preferences" since the publication of the first *Handbook of Experimental Economics* (1995). This literature's daunting size poses serious problems in terms of developing a survey since it is necessary to ignore (or only mention in passing) many worthwhile experiments, along with the flood of results that will no doubt be published shortly after this survey is completed.[1] The literature has also yielded a number of theoretical models designed to organize the data – a search for meaning based on the "facts"[2] – making this an area of experimental research where theories flow directly from the experimental outcomes (as opposed to the more usual case of experiments designed to test extant theory).

As such one must choose a point of attack to get through the literature – should it be theory or data driven? The one adopted here is "historical," using the results of a series of experiments conducted by different groups, often designed to test the latest theories used to explain earlier data. We start with a brief review of where things stood at the time the first *Handbook of Experimental Economics* (1995) was published. We then introduce the two theory papers which have had an enormous influence on this literature, Bolton and Ockenfels (2000; BO) and Fehr and Schmidt (1999; FS). These papers showed how other regarding preferences over income inequality could explain a large number of experimental outcomes, usually in small group bargaining type environments, which the "standard" economic model of strictly selfish preferences failed to organize. In contrast, the same preferences, under different institutions (e.g., competitive markets) produced the standard results. All of this was done without the need to ignore too many "dead rats" (extant results that contradict one or the other of the two models). This led to a burst of new experiments designed to distinguish between concerns for income inequality which the BO and FS models focused on and other issues such as intentionality and efficiency. We then review some of newer models designed to incorporate these experimental findings, as well experiments responding to these newer theories (e.g., Charness and Rabin, 2002). Much of the focus here will be on bargaining games (especially the ultimatum game)

[1] Our apologies in advance for those papers we have neglected.

[2] See Roth (1995) for a proposed classification of experiments according to how they were motivated and to whom they were intended to be persuasive. Under this taxonomy, searching for meaning becomes possible as results of experiments dealing with the effects of variables existing theory has little to say about accumulate, in efforts to explain the observed behavior. One of the interesting aspects of these newer theoretical models is the extent to which pure theorists have become involved in this work.

and the dictator game. While most of the literature has focused on models which rely on modifying players' utility functions to explain other-regarding behavior, there is also a strand that uses adaptive learning models for this purpose. We briefly digress to describe these papers since they not only provide an alternative explanation for some of the experimental results but also helped to spur the experimental learning literature (see Chapter xx on learning). We devote a separate section to "gift exchange" experiments, both because they have a different structure from bargaining and ultimatum games and have recently been the subject of heated debate. It should be clear by the time we finish this survey that there is no single, all encompassing model able to consistently explain all of the experimental results relating to other regarding preferences, and that a *tractable* model of this sort is unlikely to emerge any time soon.

At this point in time there are a number of surveys dealing with the other regarding preference literature that the interested reader might wish to consult. Fehr and Schmidt (2006) and Camerer (2003) cover much of the experimental literature up to their point of publication. Rotemberg (2006) surveys reciprocity and altruism in the workplace (field data), results of which are particularly relevant to the gift exchange literature.

I. Where Things Stood Circa 1995.[3]

Much of the work on other regarding preferences in 1995 hinged on results from ultimatum and dictator games. In the ultimatum game two players, 1 and 2, must decide how to divide a sum of money, *k*, between them. Player 1 (the Proposer) makes an offer to player 2 (the Responder), which if accepted is divided as player 1 proposes. However, if player 2 rejects the offer, both players get nothing. Although there are many Nash equilibria in this game, the subgame perfect equilibrium outcome in which player 1 offers the minimum amount of money required (or a small positive amount in case of no minimum requirement) is a natural equilibrium refinement under the "standard" assumption that players only care about own income. In contrast to this prediction, Proposers in developed economies typically offer between 40-50% of the pie, which Responders accept. Smaller shares are usually rejected with sufficient regularity that Proposers' income maximizing offer is in the neighborhood of 40-50%.[4] The beauty of this sequential bargaining game is that you get to see the Responder's choice in every game, and

[3] See Roth (1995) for a detailed survey of results up to this point in time.

[4] The typical experimental design is to play the game 10 times with different partners but with roles fixed. In this case one game, chosen at random at the end of the sequence, is selected for payment. Subjects also receive a small show-up fee. For an interesting cross-cultural study of outcomes for ultimatum games in a number of "primitive" cultures see Heinrich et al., 2001, 2005).

Responders face no game theoretic issues, such as ability to do backward induction or concerns about strategic uncertainty.

These initial experiments involved relatively small sums of money - $10-$30. Left unresolved was the issue of robustness, as one might suspect that with larger sums at stake, the *amount* that Responders would require to accept a proposal will go up, but the *percentage* of the pie required might well go down. There are effects of this sort, but the basic outcome, splits deviating substantially from the subgame perfect equilibrium outcome, stand. For example, Slonim and Roth (1998) conducted an experiment in the Slovak Republic where modest stakes (by American standards) had large purchasing power. They compared games in which the amount of money at stake in terms of local purchasing power equaled $30 in terms of US purchasing power, one-month's average wages in the Slovak Republic, or three-month's average wages. Their data show rejection rates decreasing somewhat as the amount of money at stake increased (from 26% under the smallest stakes to 14% with the largest, pooling over all offers strictly less than half the pie). Learning by Proposers was more rapid in games with high stakes – subjects who initially made high offers reduced their offers more rapidly while those who initially made low offers raised them more quickly than in their other treatments. Nevertheless the mean share of the money offered changed very little, averaging 45% of the pie with the smallest stakes to 43% of the pie with the largest stakes, with median offers staying in the 41-50% range throughout. Cameron (1999) reports similar results from Indonesia with the amount of money at stake ranging from a day's wages to a month's wages. She found no significant change in Proposer's behavior with increased stakes, but observes lower rejection rates in the higher stakes treatments. Controlling for differences in the distribution of offers, rejection rates were estimated to be 17% lower in the highest stakes treatment than in the lowest stakes treatment. This experiment ran for only two rounds, so there is not much that can be said about learning.[5]

One of the key questions these ultimatum game results left open was whether the close to equal splits offered were a result of Proposers "trying to be fair to Responders" or were strategic responses to anticipated rejections of low offers – the expected payoff maximizing offer in ultimatum game experiments is typically around 40 - 45% of the pie. To sort out between these

[5] See Camerer (2003) for a rather exhaustive review of ultimatum game studies.

two alternatives, Forsythe et al. (1994) compared ultimatum and "dictator" games. Like the ultimatum game, the dictator game is a two player game in which player 1 (the dictator) proposes to split a fixed sum of money with player 2. However, unlike the ultimatum game, in the dictator game player 1's proposal is binding, with player 2 having no say in the matter, as both players receive whatever the dictator proposes. This eliminates any strategic considerations from the dictator's offer, and resulted in a dramatic downward shift in offers compared to the ultimatum game: The modal offer changed from a 50-50 split in the ultimatum to game, to a zero offer in the dictator game (see Figure 1). None the less, offers in the dictator game were not all zero (or close to it) as one would expect if own income was all that mattered for dictators, and there was a cluster of equal splits.[6] The contrast between dictator and ultimatum game results clearly indicate that strategic considerations (anticipation of rejection of low offers) underlies the near equal splits typically reported in ultimatum games. At the same time they suggest some concern for the well being of others. There have since been a large number of dictator type experiments designed to sort out between various hypotheses concerning the nature of subjects' other regarding preferences. These are discussed in Section III.F along with experiments demonstrating the sensitivity of the experimental outcomes to rather modest changes in experimental procedures.

[Insert Figure 1]

Just as it wasn't initially clear whether Proposers' behavior in the ultimatum game was driven by distributional or strategic concerns, it also wasn't obvious whether the rejection of positive offers was due to purely outcomes or reflected a desire to punish unkind actions by Proposers. Blount (1995) was one of the first to show that intentions matter as she compared ultimatum games in which human Proposers made offers to games in which it was common knowledge that the proposals were generated (i) by a computer and (ii) by a "disinterested" third party. Using the strategy method, she elicited minimum acceptable offers (MAOs) from all subjects prior to their knowing if they would be randomly assigned to the role of Proposer or Responder. Figure 2 reports her results, which yield statistically significant differences between

[6] These results have since been replicated a number of times and make for an interesting classroom exercise for teaching undergraduates.

the random treatment and either the standard ("interested party") treatment or the third-party treatment, but no significant differences between the latter two treatments.[7]

[Insert Figure 2]

An interesting sidelight to this paper consists of a second treatment in which she repeats the exercise under conditions where (i) subjects knew they would be assigned to the role of Responder and (ii) they were shown the distribution from which offers will be drawn. In this treatment there was a large, statistically significant, increase in the frequency with which subjects were willing to accept the lowest possible offers ($1 or less) in the "interested party" treatment.[8] Blount attributes this difference to the fact that subjects knew their role prior to deciding, and that "…the proposal was contained in an envelope attached to their packet of materials, which led them to reason through the problem in a slightly different manner taking a much more directly self-interested approach." (Blount, p. 138).

II. Explicit Models of Other Regarding Preferences

A. *Outcome-Based Social Preference Models*

The pioneering work of Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) focused on models of players concerns about the *distribution of payoffs* along with their own income to help explain ultimatum and dictator game outcomes.[9] The real beauty of these two papers is that through simply adding concerns about the distribution of payoffs to standard concerns about own income, they were not only able to make sense of dictator and ultimatum game outcomes in which standard "selfish" economic man fails to be revealed, but also showed, without changing the structure of preferences, that standard own income maximizing results emerged in different environments. All of this was done while not needing to ignore much, if anything, in the way of inconsistent results. Both sets of authors fully recognized that other "fairness" considerations, particularly intentionality and reciprocity, were likely to play a role in experimental outcomes, but limited themselves to features needed to organize the main stylized facts at the time. In short, what these two papers did was to summarize the emerging other regarding behavior results up to that point in time, including results from gift exchange experiments, and by explicitly

[7] All of Blount's experiments involved playing a single ultimatum game.
[8] The frequency of MAOs of $1 or less went from somewhat less than 35% to over 60% judging from her figures. However, there continue to be significantly lower MAOs between the random and interested party treatments.
[9] Bolton (1991) is a notable precursor to these papers.

modeling this behavior, set off a whole new round of experiments that have helped to clarify the nature of these other regarding preferences.

The Fehr-Schmidt (FS) and Bolton and Ockenfels (BO) models assume that the utility $u_i(x)$ of an outcome $x = (x_1, \ldots, x_i, \ldots, x_n)$ for a player in the game depends on player i's own payoff x_i as well as how it compares to other players' payoffs. Some percentage of individuals in the population are assumed to get negative utility from having lower payoffs than others, which explains why Responders are willing to reject low, but positive income offers in the ultimatum game. Once Proposers recognize this, they respond by making substantial positive offers as observed in the data. There is also a portion of the population that gets negative utility from being better off than others, which can explain the positive offers in the dictator game. Both models assume that the disutility from being worse off than others is likely to greater than the disutility from being better off. A nice feature of both models is that heterogeneity is explicitly accounted for by assuming a distribution of preferences in the population. Even with a majority of players having standard preferences - concern for own income only – both models can explain the results from ultimatum and dictator games. Both models are fairly tractable since players' preferences depend only on the outcomes of the game, and not on how they have been achieved. It is therefore easy to apply both models and to make predictions about new games.

An interesting feature of both models is that they implicitly ignore payoffs outside the laboratory. This amounts to an assumption that wealth outside the lab is the same between subjects or is irrelevant to decisions made inside the lab. This assumption may seem innocuous, but may be important for experiments that study interactions between differing ethnic groups (e.g. Ferchtman and Gneezy, 2001).[10]

Formally, both models assign utility based on a subject's own payoff and an other-regarding component that compares a subject's payoff with the payoffs of others. In the FS model, this social comparison function is based on the difference between subjects' own payoff x_i and the payoffs of all other subjects in the game. Utility is reduced when i's payoff is either higher or lower than other subjects' payoffs, with the reduction being greater in the second case. The resulting utility function is shown in (1). The parameters α_i and β_i capture the marginal disutility from disadvantageous and advantageous inequality – by assumption $\alpha_i \geq \beta_i \geq 0$. It is important to

[10] See Armantier (2006) for one of the few experiments looking a the impact of wealth differences on play in ultimatum games (in this case, within experiment induced differences in player's wealth).

note that the summations in (1) cannot be replaced with the average payoff to others, as the distribution of payoffs over others affects the utility function. A second critical point is that the utility function is not assumed to be identical for all individuals. FS take advantage of this feature in exploring how the model can explain data from a variety of experiments. While the FS model is linear, there is no particular reason it cannot be modified to be non-linear.

$$(1)\quad u_i(x_i, x_{-i}) = x_i - \alpha_i \left(\frac{1}{n-1}\right)\sum_{j \neq i} max|x_j - x_i, 0| - \beta_i \left(\frac{1}{n-1}\right)\sum_{j \neq i} max|x_i - x_j, 0|$$

In the BO model, the social comparison function is based on the proportion of total payoffs a player receives. Holding own payoff fixed, utility is maximized when an individual's payoff is equal to the average payoff over all individuals. The functional form of the utility function is shown in (2). Note that the model assumes that all payoffs are non-negative. Unlike FS's model, the BO model explicitly allows for non-linear preferences.

$$(2)\quad u_i(x_i, x_{-i}) = v(x_i, \sigma(x_i, x_{-i})) \text{ where}$$

$$\sigma(x_i, x_{-i}) = \frac{x_i}{\sum_{j=1}^{n} x_j} \; if \; \sum_{j=1}^{n} x_j > 0; \; \frac{1}{n} if \; \sum_{j=1}^{n} x_j = 0$$

From a practical point of view, the two models make similar predictions in spite of their differing functional forms. The FS model can be sensitive to changes in the distribution of payoffs over other individuals (as opposed to changes in the average payoff of other individuals) which do not affect predicted behavior under BO. Consider a distribution of strictly positive payoffs over three players, P1, P2, and P3. Now imagine that a constant k, where $0 < k <$ min[P1,P2,P3], is added to P3's payoff and subtracted from P1's payoff. Under BO, this cannot impact P2's utility as his payoff share is unaffected. With the FS model, this will lower (raise) P2's utility if the original payoffs are strictly increasing (decreasing) in the player indices. The BO and FS model also make slightly differing predictions as more players are added to the game, an issue which we discuss below in the context of three person ultimatum games.

Remarks: The BO and particularly the FS models have been met with a number of criticisms. We leave the issue of intentions and reciprocity aside for the moment. BO and FS both made it clear that they understood these factors to be present. Their models are not intended to be complete theories of all factors involved in other-regarding behavior, and therefore do not incorporate features that are unnecessary to rationalize the existing data. Experiments designed to better understand the role of intentions and reciprocity are discussed in detail in the next section.

Rotemberg (2008) notes that both BO and FS have trouble explaining the common use of even splits in the ultimatum game. The point is actually easier to see in terms of dictator games, which also have an atom at the even split. BO predict that there should be no even splits in the

dictator game since the marginal utility from own income is positive at an even split and the marginal utility from the other regarding component of the utility function is zero. FS rationalize even splits in the dictator game (as well as the ultimatum game) in terms of $\beta > \frac{1}{2}$ so that individuals strictly prefer a dollar in the pocket of someone with lower income then themselves to a dollar in their own pocket. Rotemberg argues that this represents an implausibly high level of altruism, and also notes that introducing a non-linear utility function in place of the piecewise linear function does not improve matters. He then offers a model in which Responders treat low offers as a signal of strong selfishness on the part of Proposers that they want to punish. That is, like FS and BO Rotemberg explains rejections of unequal splits in the ultimatum game in terms of ill-will toward Proposers, but argues that this the ill-will is not a result of income inequality.

Shaked (2006) offers a much more sweeping indictment of the FS model. The argument can be summarized as follows: By virtue of having an infinite number of possible parameters values the theory can predict a wide range of outcomes, from the competitive to the cooperative, so that its predictions depend on the value of these parameters. (The infinite number of parameters values referred to concern the *heterogeneity* of preferences within any given sample population so that for any given game the theory's predictions depend on how inequity averse the population is.) As such the theory has no explanatory value beyond the capacity to predict a broad range of outcomes as a function of possible parameter values within a given population.

We agree that there clearly are problems with both the BO and FS models, but holding these models to their point predictions is probably too stringent a standard. Even in settings where intentions and reciprocity have little force, any attempt to have a single functional form fit the infinite variety of preferences present in the population is bound to lead to some questionable results. The mass at the even split in dictator game makes this point clear. As Andreoni and Bernheim (2008) argue persuasively, these individuals are probably following a social norm with the goal of appearing "fair" rather than maximizing a well-behaved utility function over the distribution of payoffs. In other words, the behavior of these subjects is driven by objectives outside the realm of *any* model of purely outcome based preferences. What BO and FS did quite successfully is to provide a tractable model that can rationalize and synthesize a reasonably large body of data. Their work also motivated a large number of new experiments designed to better understand what drives deviations from the standard (strictly) selfish preference model.

III. Experiments Responding to Theories of Other-Regarding Preferences

A. *Some Initial Tests of the Bolton-Ockenfels and Fehr-Schmidt Models*

Initial tests of the BO and FS models focused on two issues: (1) the extent to which choices depend not only on outcomes but also on how those outcomes were achieved (i. e., the extent do reciprocity and perceived intentions play a role in these games) and (2) the scope of players concerns for own income compared to others in the relevant reference group.

Falk, Fehr, and Fischbacher (2003; FFF) investigated the role of intentions in a series of four discrete "mini" ultimatum games in which the Proposer chose between two possible allocations *x* and *y*. In all four games the reference point allocation *x* was the same: an (8, 2) split where the Proposer's share is listed first. In one game the alternative allocation was a (5, 5) split compared to which the (8, 2) allocation is relatively selfish. In the second and third games the (8, 2) split

was paired with a (10, 0) division and a (2, 8) division respectively. Compared with a (10, 0) option the (8, 2) split is relatively fair, with the (2, 8) division forcing the Proposer to chose between being fair to himself or to the Responder. Finally, as a control treatment they paired the (8, 2) allocation with itself, so that the Proposer had no choice but to offer an (8, 2) allocation.[11]

Subjects played all four games in different order, with no feedback following each game. The strategy method was employed so that Responders had to indicate their choices for both the reference point allocation and the alternative allocation. Figure 3 reports their results in terms of rejection rates for the (8, 2) allocation in favor of zero payoff for both players. The rejection rate in the game with the (5, 5) alternative is significantly higher than in all the other games, with the difference between the (2, 8) and (10, 0) games statistically significant as well. Proposer behavior anticipates these outcomes as the percentage of (8, 2) offers is 31%, 73% and 100% against the (5, 5), (2, 8) and (10, 0) alternatives respectively. FFF conclude that differences in rejection rates between treatments clearly indicate that intentions matter. Finally, there is an 18% rejection rate under for the (8, 2) allocation when the Proposer has no choice, which FFF cite as evidence of pure income inequality aversion.[12]

[Insert Figure 3]

Three person ultimatum games, introduced by Güth and van Damme (1998; GD), have also provided a productive framework for testing the BO and FS models. In GD player X proposes to split income between players X, Y and Z. Player Y then accepts or rejects the split, with the division binding when Y accepts and all three players getting zero if the proposal is rejected. Z is a "dummy" player with the same role as player 2 in the dictator game. GD found that Proposers (player X) took advantage of Z's Dummystatus, essentially dividing the money between themselves and Y, with these offers rarely rejected when Y had full information about the proposed split.

Bolton and Ockenfels (1998) cite these results as strikingly consistent with their model. In the BO model, the other-regarding component of utility is evaluated relative to the social norm of equal shares for all players. Accordingly, adding a third player to the ultimatum game changes the equal division social norm from ½ to $\frac{1}{3}$, leading to a prediction of higher acceptance rates

[11] Bolton and Zwick (1995) are responsible for introducing the mini (or cardinal) ultimatum game to the experimental literature.

[12] See Brandts and Solà (2001) for a similar design and results. These papers should be regarded as independent and simultaneous.

for offers in the interval $[⅓, ½]$ than in the typical two player game. In line with the prediction, GD observed no rejections of offers in the neighborhood of 40% of the pie for Responders (player Y) when they knew the full distribution of payoffs. In contrast, such offers have a rejection rate of about 20% in a standard two-person ultimatum game (see Cooper and Dutcher, 2009). Moreover, other-regarding preferences in BO only depend on *own* share of total payoffs. The distribution of payoffs over the other players has no impact on utility. This lines up well with GD's observation that no rejections could be attributed to the low share allocated to the Dummy player.[13]

Kagel and Wolfe (2001; KW) modified GD's three player design to obtain a much more demanding test of the BO and FS models. First, the responding player was randomly selected to be either Y or Z *after* X had made her allocation. This was designed to maximize the chance of the Responder getting a relatively low offer as Proposers, not knowing the identity of the player prior to making an offer, could no longer pay off the Responder at the expense of the Dummyplayer. Second, the rejection outcome for the non-responding player varied between treatments taking on values of \$0, \$1, \$3, and \$12 with the amount of money to be divided set at \$15. Given a positive consolation prize for the "dummy" player, the BO model predicts that the Responder will accept *any positive offer* since if they reject it they get no money and earn less than the average payoff. The FS model permits some rejections with a positive consolation prize. However, once the level of disadvantageous inequality from rejection is greater than from acceptance, the offer must be accepted regardless of the amount of advantageous inequality the offer provides the Responder compared to the "dummy" player.[14] As such the \$12 consolation prize treatment should effectively eliminate all rejections under the FS model as well.

Pooled Responder data from KW's treatments with positive consolation prizes are reported in Figure 4.[15] The point prediction of the BO model is falsified as rejection rates average between 15%-22% under the different positive consolation prize treatments. Further, with the \$12 consolation prize, virtually all offers should have been accepted according to the FS model, but

[13] The FS model has more difficulty characterizing the GD results.

[14] Consider the following proposal (9, 6, 0) where 9 is the Proposer's share, 6 is the Responder's share and the Dummy gets 0. In case of rejection both the Proposer and Responder get 0 and the Dummy gets some amount greater than 3 (3+). As such utility from acceptance is $U_i = 6 - 0.5\ \alpha_i\ (3) - 0.5\ \beta_i\ (6)$ and utility from rejection is $U_i = 0 - 0.5\ \alpha_i\ (3+) - 0.5\ \beta_i\ (0)$ so that with $0 \leq \beta \leq 1$, as FS assume, acceptance dominates rejection.

[15] KW used a between groups design with 10 rounds per treatment with one of the 10 selected at random to be paid off on. Subjects received feedback following each round.

this treatment had the highest rejection rate of 20%. Even more damaging to both models, the rejection rate for the $0 consolation prize treatment was 21%, falling in the same range as rejection rates for the positive consolation prize treatments. Similar results were found with a *negative* consolation prize of $10 for the Dummy player.[16] In short, the consequences for the Dummy player did not seem to matter to Responders.[17]

[Insert Figure 4]

The invisibility of the third player in KW's experiment calls into question FS and BO's explanations of why, despite other regarding preferences of the sort both models specify, there is typically no impact on the standard (selfish) model's predictions in competitive markets. For both models, explaining market results such as those reported by Roth et. al (1991) relies on individuals comparing their payoffs with those of ***all*** market participants. For example, FS note that the crucial factor leading to very inequitable outcomes in market games is that no single player can enforce an equitable outcome. Therefore, even very inequity-averse Responders try to turn the unavoidable inequality to their advantage by accepting low offers. However, in KW's three player ultimatum game we see a relatively high percentage of Responders who get low offers turning their backs on the opportunity to reduce their income inequality relative to the Dummy player by accepting a modest positive offer. Another key difference between the three player ultimatum game and market games is that small payoffs can be directly attributable to a single person in the ultimatum game, whereas individual attribution is typically difficult in market games. As such the results also suggest that intentions matter, but with the added twist that "unfair" offers are rejected regardless of the consequences for "innocent" third parties.

Bereby-Meyer and Niederle (2005; BMN) report two three-person games designed to distinguish the presence of outcome based preferences and reciprocity in bargaining games. The first class of games (called third-party rejection payoff games – TRP) is similar to the three player game in KW with a Proposer making an offer to a Responder under each of three consolation prizes for the third "dummy" player - $0, $5, or $10. There are a number of procedural differences from KW as each subject plays once under each treatment with no feedback on outcomes until the session is over, and the Proposer chooses to split the money

[16] In this treatment all subjects received a starting cash balance of $15 in place of the $5 show-up fee provided in the other treatments in order provide positive earnings for Dummy players in the case of rejections.

[17] Random effect probits indicate that, controlling for offers, rejection rates were essentially unaffected by the presence of positive consolation prizes or the size of the consolation prize.

($10) strictly between herself and the Responder. In the second class of games (referred to as proposer rejection payoff games – PRP) the Proposer is required to split the $10 between the Responder and the Dummy player with no money for herself. If the Responder accepts the division is binding. If she rejects both she and the Dummy get nothing, but the Proposer gets a rejection payoff - $0, $5, or $10 - depending on the treatment. In terms of pure intention based models,[18] in the PRP-$0 game the Responder should accept all offers since the Proposer's payoff does not depend on the Responder's action and the Dummy, not taking any action, cannot signal kindness one way or the other. In the PRP-$5 and $10 games pure intention based models allow for the possibility that low offers that would be rejected in the parallel TRP games will be accepted so as to not reward Proposers for unkind behavior.

[Insert Figure 5]

Figure 5 reports their results (all proposals were in dollar increments with $1 being the smallest possible allocation). In the TRP games with positive payoffs for the Dummy player when offers are rejected, low offers are routinely rejected at the same, or higher, rates compared to the TRP-$0 treatment, which is inconsistent with both BO and FS models. Even more damaging, there were significantly higher rejection rates in the TRP game than in the PRP game for all payoff levels, which is inconsistent with outcome based models that predict no difference in behavior according to players' positions in the allocation process (Proposer, Responder or Dummy player). Pure intentionality models can explain why rejection rates for low offers were significantly lower in the PRP-$5 and $10 games, but cannot rationalize the higher rejection rate observed in the PRP-$0 games when Responder's payoffs were $3 or less. The results of this experiment point to multiple forces playing a role in other-regarding behavior. The data give the impression that intentions play a larger role than purely distributional concerns, but the role of pure outcome based preferences is far from zero.

Xiao and Houser (2005; XH) also report results from a one-shot ultimatum game that are quite damaging to the FS and BO models. They add the interesting twist of Responders having the option to send written messages to Proposers in addition to deciding whether to accept or

[18] It is a little tricky to define a "pure" model of intentions, since actions are defined as kind or unkind in terms of the distributions of payoffs that result. An axiom like the following captures the spirit of a pure model of intentions reasonably well: Consider a binary choice between two distributions of monetary payoffs across individuals. If these distributions result solely from moves by nature (in the game theoretic sense) or choices of disinterested parties (individuals whose monetary payoffs do not depend on their choices or the choices of other individuals), then all individuals should prefer the distribution that gives them the higher monetary payoff. In the PRP-$0 game, the Proposer is a disinterested party and the Dummy plays no role in determining the distribution of payoffs.

reject offers. XH find that conditional on offers being 20% of the pie or less, rejection rates drop from 60% to 32% when Responders have the option to verbally punish Proposers – for about 80% of the low offers a message with "negative emotions" was sent. There were no significant differences in rejection rates with and without communication for more generous offers, as well as no significant differences in the distribution of offers.

XH's results are best understood in terms of the costs of punishment. Responders have an emotional reaction to low offers and reciprocate by punishing the Proposers. When the *only* punishment mechanism available is the relatively costly option of rejection, they use it. However, as demonstrated by Andreoni and Miller (2002), other regarding behavior is price sensitive. Given the less costly option of punishing selfish Proposers with verbal abuse a number of Responders chose it rather than give up the money. Thus, outcome based models like the BO and FS that only consider pecuniary outcomes are likely to miss important aspects of subjects' behavior when players have a wider array of options to consider.

An important question here is the long run implication of Responders using verbal rather than monetary punishment. In another paper, Xiao and Houser (2008) argue that monetary punishment is initially more powerful than verbal punishment. One implication of this is that if Responder's persistently eschew monetary punishments, this would result imply a long run trend towards lower offers.[19] However, there is no reason to believe that Responder's mix of punishments will be constant over time. Additional experimental work is needed to determine the long run equilibrium when both pecuniary and non-pecuniary punishments are available.

Remark: This experiment holds lessons for field studies of other-regarding behavior. These studies typically focus on the same narrow avenues of response incorporated into laboratory experiments. But in field settings subjects usually have an array of responses available to them that are difficult to capture or quantify. Failing to identify a reciprocal response along one avenue doesn't mean that a reciprocal response hasn't occurred, nor does it imply that reciprocity isn't playing an important role in driving individual's choices. As such one needs to be particularly careful when drawing conclusions from less structured field settings.

Summary: The BO and FS outcome based models of social preferences pulled together a surprisingly large number of experimental outcomes and organized them using other regarding preferences based on income inequality. As such they provided a clear focal point for further

[19] In this follow-up paper XH look at the effect of Responder messages in a one-shot dictator game, finding that it results in higher offers.

experimental work. Although subsequent experiments have been hard on both models, these papers made (and continue to have) a major impact on the literature and have moved the discussion forward in terms of helping to identify exactly what kind of fairness considerations underlie deviations from the standard selfish, income maximizing model. At this point the data suggest: (1) outcomes do matter, to some extent at least, as for example when Responders reject unequal offers in cases where Proposers have no choice but to make such offers, and (2) intentions matter as well, possibly even more than outcomes.

There still remain important methodological issues to be addressed in this literature. Reading through all of these papers at once, we are struck by the varying methods used by different researchers. Strict comparisons across papers are therefore difficult, and it is unknown how many of the results reported in the literature are robust to changes in methodology. As a vivid example of this problem, consider the following question: why do subjects in FFF reject unequal offers when Proposers have no choice but to make such offers? Difference aversion is an obvious explanation for this, but results from Charness and Rabin (2002; CR) force us to question this. In a dictator game, CR ask player Bs to choose between (800, 200) – 200 being B's payoff and 800 being A's payoff – versus a (0, 0) allocation and find that 100% of the 36 subjects queried chose the (800, 200) option. It is hard to argue that the differing results of FFF and CR are anything other than an artifact of how preferences are being elicited. One possible methodological cause is CR's use of an "equal opportunity" procedure whereby each subject got to choose as a B player knowing that their actual position as the A or B player would be determined randomly at the end of the session; other researchers have found that "equal opportunity" procedures reduce inequality aversion (Bolton and Ockenfels, 2006). Another possibility is that subjects don't fully understand that the Proposer's choice is irrelevant in FFF, but would understand this if they gained experience via repeated trials with feedback. The point is that further work is needed to know how researchers' differing methodological choices are affecting the observed behavior and, by extension, conclusions reached with respect to other-regarding preferences.[20]

The question of whether to use one-shot experiments (or repeated trials without feedback) versus repeated trials with feedback comes up repeatedly in this literature. The argument for

[20] For another clean example where a framing effect changes the degree of other-regarding behavior, in this case a preference for inclusion, see Cooper and Van Huyck (2003).

using one shot experiments is that these are particularly clean – there is no possibility of unwanted repeated game effects and, because there is no rematching of subjects, games from the same session can be treated as fully independent observations.[21] However, the decision to not allow for learning via experience can affect results. For example, if one looks at a standard ultimatum game played for ten rounds, the distribution of proposals is typically much more dispersed in early compared to later rounds, with the frequently stated stylized result of a high concentration of offers in the 40-50% range (and minimal rejection rates for such offers) only emerging in later rounds (see, for example, the data reported in Roth et al., 1991).

Economists have traditionally preferred experiments with repeated trials and feedback in response to the original Wallis-Friedman (1942) critique of economic experiments:

> "It is questionable whether a subject in so artificial an experimental situation could know what he would make in an economic situation; not knowing, it is almost inevitable that he would, in entire good faith, systematize his answers in such a way as to produce plausible but spurious results.
>
> For a satisfactory experiment it is essential that the subject give actual reactions to actual stimuli.... Questionnaire or other devices based on conjectural responses to hypothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react."

This is not to say that repeated trials are the only way to conduct experiments, but more investigation is needed of whether results based on one-shot experiments, or experiments without feedback, yield results that are robust to subjects gaining experience.

Another methodological question that comes up frequently in this literature is whether to use the standard direct response method or the strategy method. The appeal of the strategy method is obvious, as it allows more data to be gathered per subject, but once again the question comes up of whether this affects behavior. Results vary as to whether the strategy method leads to different results than direct response.[22] At the very least it seems clear that it *can* matter. Brandts and Charness (2003) and Brosig, Weimann, and Yang (2003) both find that punishment rates for an unkind and/or deceptive act are significantly lower when the strategy method is used. Along similar lines, Casari and Cason (2009) find significantly less trustworthy behavior in trust

[21] This depends a little on how careful you want to be in running statistics. There are possible sources of session effects beyond direct interaction (i.e. the instructions are read slightly differently, sunny vs. rainy weather affects the mood of subjects, etc.). Frechette (2007) provides a good discussion on how to control for session effect in experimental data.

[22] See Brandts and Charness (2009) for a recent survey on this issue.

games when the strategy method is used. The size of the effect can be large. In Brandts and Charness, the clearest example of an unkind act occurs when a player lies about his intent to make a fair choice. Using direct responses this sort of lying is punished in 56% of the observations (9/16), but the punishment rate is halved to 28% (19/69) when the strategy method is used. Casari and Cason observe that 40% of subjects (14/35) return nothing when the direct response method is used, but this jumps to 60% (43/72) when the strategy method is used. In contrast to the preceding, there are also cases where the strategy method does not matter, as in the examples reported by Brandts and Charness (2000). When an effect exists, the strategy method yields less reciprocal behavior than direct responses. This suggests an anchoring and adjustment process in line with other examples of framing effects – when subjects are faced with a problem that has multiple dimensions, the framing can impact which dimension gets the most attention and which is treated as a secondary concern.[23] In the environments discussed above, subjects are trading off reciprocity for kind/unkind actions against payoff maximization. Direct responses seem to focus attention more on reciprocity, yielding more reciprocal behavior. From a practical point of view, the issue is that other-regarding behavior can be sensitive to the method of elicitation. This makes it difficult to directly compare studies which have used different elicitation methods.

B. *Social Preferences versus Difference Aversion*

At this point in time it seems safe to state that reciprocity, or intentions, play an important role when behavior deviates from the standard selfish model. But many observers still see a role for outcome based preferences on choices. Within the BO and FS models, difference aversion is the driving force behind these outcome based preferences. Beginning with Charness and Rabin (2002) and continuing with Engelmann and Strobel (2004, ES) it has been argued that social welfare preferences – concerns for efficiency (defined as maximizing total payoffs for the group) and the payoffs for the least well off players in the group (maximin preferences) – are the key factors underlying outcome based preferences rather than difference aversion. Both of these papers report a number of results supporting this position.

For example, consider player 2 choosing over distributions A, B, and C in cases X and Y shown below. For X player 2's payoff is independent of their choice.

[23]See, for example, Tversky, Sattath, and Slovic (1988).

Choice X				Choice Y		
Allocation	A	B	C	A	B	C
Player 1	16	13	10	16	13	10
Player 2	8	8	8	7	8	9
Player 3	5	3	1	5	3	1
Total[a]	29	24	19	28	24	18
Percentage[b]	70.0	26.6	3.3	76.7	13.3	10.0

[a] Sum of players' payoffs. [b] Frequency with which the allocations were chosen in ES (2004)

Alternative A is efficient, maximizing total payoffs. It is also a maximin allocation as it has the highest payoff for the least well off player. In contrast, B maximizes players 2's utility according to the BO model, with C maximizing player 2's utility according to FS.[24] The allocations in choice Y are the same as X with the exception that in Y it costs player 2 a modest amount of money to choose the efficient, as well as the maximin, outcome. BO now predicts choice of B or C, with FS still predicting C.[25] In practice this small increase in cost has little impact on the choice of the efficient (as well as maximin) allocation. Looking at a variety of choices of this sort, both CR and ES estimate the relative impact of efficiency considerations, maximin preferences and difference aversion of the sort specified in BO and FS on choices, concluding that social welfare preferences play a more important role than difference aversion.

Responding to ES, BO (2006) argue that the essential question is the willingness to pay for efficiency as opposed to equity. They provide results from an experiment in which twice as many subjects deviate from higher own payoffs in favor of the more equitable outcome as opposed to deviating in favor of the more efficient outcome. In response ES (2006) point out that it is difficult to identify the correct metric for measuring the tradeoff between efficiency and equity, noting that in BO's experiment subjects are asked to pay a lot for relatively small percentage increases in efficiency. FS's (2006) response to ES is to identify strong subject population effects in the degree with which subjects favor efficiency over equity. They replicate one of ES choices, reporting that 53% of non-economists prefer the most egalitarian (and least efficient) allocation as opposed to 30% when the subjects are economics and business students.

[24] For BO it is own share divided by the average share that determines σ_i, which is maximized for allocation B. For FS the relatively small difference with respect to 2's share relative to 1 in allocation C tips the scales given the other differences and the greater weight placed on negative as opposed to positive differences FS assume.
[25] Payoffs were in Deutch Marks with an exchange rate of between $0.45 and $0.55 at the time of the experiment.

They attribute the stronger preference for efficiency over equity in ES's experiment as opposed to other studies to the fact that ES's subject population consisted of economics and business majors.

Reviewing this exchange of views, the literature suffers from attempts to oversimplify subjects' behavior. Not only is there great heterogeneity among subjects' preferences, as FS (2006) convincingly demonstrate, but subjects also appear to be able to make reasonable adjustments in how much they rely on any one criterion depending on the costs and benefits involved. Further, reciprocity/intentionality appears to be a stronger force in driving subjects' choices than any purely distributional concerns: Recall that only 18% of Responders rejected the (8, 2) offer in FFF when Proposers have no choice. In contrast, 72% of such offers are rejected in a standard ultimatum game when Proposers could chose between allocations (Cooper et al., 2003).

C. Results from Multilateral Bargaining Experiments

Multilateral bargaining experiments provide a different window into many of the other regarding preference issues discussed so far. In multilateral bargaining games a set of n players must decide on an allocation of a sum of money k through a voting mechanism. In the simplest set-up, all players make proposed allocations, one of which is selected at random to be voted on under majority rule with no opportunity to amend proposals.[26] Consider an infinite horizon version of the game so that if a proposal is rejected new proposals are solicited and the process repeats itself until an allocation has been made. The money, k, is reduced to δk following each round in which a proposal is rejected, $0 < \delta \leq 1$. As with bilateral bargaining games any proposal that is accepted constitutes a Nash equilibrium. Similar to subgame perfection, the preferred equilibrium refinement is that of a stationary subgame perfect equilibrium (SSPE) which, roughly speaking, is a subgame perfect equilibrium in which the history of past choices plays no role in proposals or in voting (Baron and Ferejohn , 1989).

A shrinking pie ($\delta < 1$) is *not* necessary to obtain an equilibrium in these games. A core element of the SSPE is formation of a minimum winning coalition (MWC) in which the Proposer gives payoffs to just enough players to secure passage of a proposal, and zero to everyone else. The threat of being left out of the money in case of rejection induces players in the MWC to vote

[26] This model was originally developed, and has been used extensively, to provide a game theoretic framework for legislative bargaining (see Palfrey, Chapter xx, for a review of legislative bargaining games).

in favor of the proposal provided they have been given a sufficiently high payoff. Thus, with $\delta = 1$, *there are no efficiency issues at stake in accepting or rejecting offers*, and the frequency of MWCs and/or the sensitivity of players to other players getting zero payoffs provide insight into maximin preferences in a bargaining environment.

Several experiments of this sort have been conducted. Frechette, Kagel and Morelli (2005a; FKM) ran five player games where each player had equal voting weight. MWCs were formed 77% of the time for inexperienced subjects, and 94% of the time for experienced subjects.[27] In equivalent three player games, MWCs averaged 69% of all proposals with this number increasing to 85% for experienced subjects (FKM, 2005b). There were very few perfectly egalitarian proposals in these games, averaging well below 10%. Further, random effect probits consistently show own payoffs to be the key factor determining whether or not to vote for a proposal, with dummy variables accounting for the number of players getting zero shares *not* achieving statistical significance in both experiments. If maximin preferences play an important role in subjects' decisions, one would expect some sensitivity to the plight of players receiving a zero allocation.

However, as is often the case, these results are not conclusive in knocking out maximin preferences. The problem is that given the systematic growth of MWCs one cannot distinguish between learning the benefits of MWCs from responding to the "selfishness" involved in other players forming MWCs (or anticipating other players' selfishness). That is, the data can potentially be accounted for by a selfish core of players in conjunction with a group of conditional reciprocators. It would be helpful to distinguish between these two alternatives.

FKM (2008) study a five player linear public goods game in which the allocation of funds to the public and private goods is decided by majority voting (as opposed to the usual voluntary contribution mechanism). Payoffs are a linear function of the amount of the budget allocated to public goods and to a player's own share of the goods, with players having homogenous preferences for public versus private goods. Treatment conditions included a variety of weights attached to public versus private goods, which had the effect of varying the marginal return to the public good, with the treatment of primary interest for present purposes

[27] In these, and the other experiments reported here, there were several groups bargaining at the same time, with subjects randomly allocated to a new group following each bargaining round with 10 or 12 rounds per session, with one round per session selected at random to determine payoffs. Data reported are for all proposals – passed or otherwise. Probits looking at voting exclude the votes of Proposer of a given allocation.

being one in which the theory predicts an all private good allocation within a MWC. The predicted allocation in this treatment not only gave zero payoff to two of the five players, but was less efficient than the perfectly egalitarian all public good allocation in terms of the total money payout to subjects ($37.50 versus $43.75).[28]

Thus, in this case both maximin preferences and efficiency considerations favor the all public good allocation. However, the data show little concern for either as all public good allocations account for only 3% of all proposals, versus 65% of all proposals involving MWCs.[29] The incentives for MWCs are clear enough in this case as Proposers averaged $15.64 for allocations that passed, with coalition partners averaging $10.93, a little over $2 more than with an all public good allocation. Both CR and ES recognize tradeoffs between social welfare preferences and own payoffs, so that the one can rationalize the high frequency of equilibrium type offers in this experiment on the basis of the higher own payoffs achieved with equilibrium type offers. However, the key point is that the efficient all public good allocation (which is also a maximin allocation) had very little drawing power at any point in these sessions, suggesting little weight placed on social welfare preferences.

There is clear evidence that differences in payoffs matter *within* MWCs in these experiments as the SSPE allocation typically calls for a much more uneven distribution of payoffs within the MWC, a distribution that is hardly ever offered and which voting regressions indicate would have virtually no chance of passing. In this respect it's worth noting that probit estimates of Responders' *average* indifference point for the share required to accept or reject a given allocation is close to what the SSPE predicts, but the dispersion in voters' preferences makes it almost certain that these allocations will not get passed. Intuitively, one might guess that the FS or BO models might be able to capture this aversion to strong payoff differences within MWCs. However, Montero (2007) shows that if players have FS type preferences, Responders should be willing to accept *smaller* shares than predicted under the SSPE.[30] As such

[28] The amount of money available to be distributed shrinks by 20% in these games if proposals are rejected.

[29] Some fraction of these MWC allocations included a small public good allocation (averaging less than 5% over all equilibrium type offers). The bulk of the remaining offers involved some private goods allocated to all players. Forty percent of these cases involved token allocations to two players (allocations that summed to less than 10% of the pie), constituting near MWC type allocations, versus 20% fully egalitarian private allocations which, in this case, were dominated by an all public good allocation (suggesting small mistakes for some players).

[30] As an extreme example, assume that disutility is only experienced from disadvantageous inequality and there is no discounting. Consider the position of Responder j who is a member of a MWC in a game with five players where all others are following the SSPE. Rejecting a proposal, j gains very little as there is a 40% chance that the outcome is unchanged (j is again a member of the MWC), a 20% chance that disadvantageous inequality is lowered (as j

the failure of allocations to approach anything close to the SSPE allocation, or below it, should be counted as further evidence that more is going on in games of this sort than we currently understand.

D. Models Incorporating Reciprocity/Intentions of Proposers

Experiments investigating the FS and BO models make it clear that a full theory of other-regarding preferences must account for Proposers' intentions as well as the payoff distribution. A number of theory papers address this issue by developing models which incorporate intentions and reciprocity.[31] While this literature addresses some delicate theoretical issues,[32] our interest is primarily in how these models enrich our understanding of experimental data. The work of Charness and Rabin (2002) has been a particularly influential example of this sort of work. They propose a simple model that allows both for a wide variety of outcome based preferences as well as reciprocity. The functional form is shown in (3). The variables ρ, σ, and θ are parameters, r and s are indicators for $x_j > x_i$ and $x_i > x_j$ respectively, and q = - 1 if the other player has "misbehaved" and equals zero otherwise. Note that the utility function gives j's utility, corresponding to the notion that the utility of a "Responder" is being measured

$$\textbf{(3)} \quad u_j(x_i, x_j) = (\rho \cdot r + \sigma \cdot s + \theta \cdot q)x_i + (1 - \rho \cdot r - \sigma \cdot s - \theta \cdot q)x_j$$

This model is intended primarily to provide intuition for understanding the experimental data and can be thought of as a reduced form model rather than a full model of other-regarding preferences. A nice feature of the model is that, ignoring reciprocity, the functional form is sufficiently flexible to nest a number of concepts about other regarding preferences such as competitive preferences, difference aversion and social welfare. By fitting the parameters of the model to the data it is possible, in principle, to sort out what elements of other-regarding preferences best explain behavior, with their experimental work consisting of an exercise of this sort. In practice, misspecification of the model, as well as heterogeneity between subjects,

---

become the Proposer and gets more money), and a 40% chance that the disadvantageous inequality becomes far worse (as j is left out of the winning coalition and gets nothing). The latter makes the Responder strictly prefer voting in favor of the offer since the expected loss in income is greater than the expected gain from being the Responder and yields more extreme disutility from disadvantageous income inequality than voting in favor of the proposal.

[31] For an important precursor to this literature, see Rabin (1993). Recent papers developing theoretical models of reciprocity and other-regarding preferences include Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), and Cox, Friedman, and Gjerstad (2007). See Cox, Friedman, and Sadiraj (2008) for an interesting axiomatic approach to other-regarding preferences.

[32] Dufwenberg and Kirchsteiger (2004) provide a good discussion of the theoretical issues.

makes econometrically distinguishing between different types of other regarding preferences an extremely difficult task.

The roll of reciprocity in the CR model lacks a solid foundation – the weight on another player's payoff is reduced if they have "misbehaved," a vague term which, at best, serves as a useful simplification for interpreting the data. Full understanding of the model requires reading the Appendix which delivers a detailed theory of how good and/or bad behavior can be endogenously defined. The critical role of beliefs is highlighted, as good or bad behavior can only be defined in relationship to what was expected. This illuminates a central problem in empirical examinations of reciprocity – while distributions of outcomes are easily observed, beliefs are not directly accessible to researchers.

E. Decomposing Behavior in the Investment/Trust Game

The investment (or trust) game introduced in Berg, Dickhaut, and McCabe (1995; BDM), is a sequential move game in which two players are given equal endowments. Player 1 moves first, with the opportunity to send money to player 2. The amount of money sent to player 2 is typically tripled, after which player 2 has an opportunity to send money back to player 1, after which the game ends. With standard selfish preferences, player 1 should anticipate player 2 not returning any money, and therefore not send any money. Experiments show player 1s sending positive amounts, with positive amounts returned, although typically (i) there is considerable variability across subjects in the amount of money sent and returned, and (ii) the average amount retuned is somewhat less than the amount of money sent.[33] The game is usually played only once in a given experimental session.

Cox (2004) conducts an experiment to begin to determine the other regarding preference factors underlying the trust game. Using a between groups design he conducts a standard investment game (the control treatment) as well as two variants of a dictator game. The first dictator game differs from the investment game in that player 2 has no decision to make as they have no opportunity to return any money. In the second dictator game, player 1s do not make any choices. Rather they are given endowments equal to the amount of money kept in the control treatment with player 2s given endowments equal to the amount of money received in the

[33] In BDM the average amount of money player 1s sent was $5.16, with the average returned equal to $4.66. This game is also sometimes referred to as the moonlighting game.

control treatment.[34] After being told the additional the additional dollars they have relative to player 1s, 2s have an opportunity to send money back to player 1s.

Any money sent in the first dictator game represents altruistic other-regarding preferences (or a taste for efficiency as the money is tripled) as distinguished from possible trust and anticipation of positive reciprocity. In turn, any money "returned" in the second dictator game represents other regarding preferences resulting either from difference aversion, or maximin preferences, as opposed to the investment game where any money returned represents reciprocity in conjunction with these other factors. As expected, the amount of money sent in the control treatment ($5.97) is greater than in the first dictator game ($3.63; Treatments A and B in Figure 6). This difference, approximately 40% more money sent in the investment game than the first dictator game, can be attributed to trust and anticipated positive reciprocity as opposed to altruism or a taste for efficiency. The amount of money returned in the second dictator game averaged $2.06 compared to $4.94 in the control treatment so that approximately 58% of the money returned in the investment game can be attributed to positive reciprocity.

[Insert Figure 6]

Thus, while trust and positive reciprocity play a role in the trust game, other forces are at work so that behavior in the trust game should be regarded as an imperfect measure of trust and trustworthiness. Consider the results of Glaeser et al (2000). They compare data from a one-shot trust game with results from surveys measuring subjects' attitudes towards trust as well as their past trusting and trustworthy behavior. Glaeser et al report a positive but far from perfect correlation between behavior between the experiment and survey results.[35] They interpret their results as illustrating the strengths and weaknesses of surveys. But a better explanation might be that there is measurement error with respect to trust and trustworthiness in *both* the survey instrument and the trust game. The positive correlation reflects a common unobserved behavioral trait that presumably coincides with trust (or trustworthiness), but the experimental data also reflect other behavioral traits such as outcome based. This suggests that a combination of surveys and experiments will be a better predictor of behavior in field settings than either instrument in isolation.

[34] To be precise, each pair in the second dictator game is matched with a pair in the control treatment. The endowments in the second dictator game match the amounts kept and received by the corresponding pair in the control treatment.

[35] Interestingly, senders' behavior is better predicted by the survey about past behavior while receivers' behavior correlates well with components of the attitudinal surveys related to trust (rather than trustworthiness).

An interesting feature of the investment game data is that player 2s typically return less money than what player 1s send (before any multiplication takes place), so that sending money is typically not profit maximizing.[36] This raises the interesting question of whether or not the amount of money sent would deteriorate over time in an experiment with repeated trials and random rematching of players.

F. A Second Look at Dictator Games

The dictator game was originally designed to distinguish whether the near equal shares Proposers offer in the ultimatum game are a result of Proposers "trying to be fair" to Responders or a strategic response to anticipated rejections of low offers. It has, however, become a very popular tool for trying to distinguish between various theories of other-regarding preferences. Dictator games have been used to calibrate the Fehr-Schmidt utility function by pinning down the relative weights put on advantageous versus disadvantageous inequality (FS, 1999; Blanco, Engelmann, and Norman, 2006), assuming (or in the case of Blanco et al, testing) that these estimates extrapolate directly to other, more complicated environments. Choices in dictator type games provide some of the strongest experimental evidence for social welfare preferences, as well as preferences for efficiency (CR, 2002; ES, 2004, 2006).

While the dictator game was well suited for its original purpose, it is unclear how much can be learned from experiments designed to precisely identify the form of outcome based preferences. Dictator games are an attractive tool for this task because of their simplicity and because neither reciprocity nor strategic uncertainty play a role in decision making. Given evidence cited elsewhere in this chapter that reciprocity and intentionality tend to be stronger forces than purely outcome based preferences, an environment where these factors play little role (like the dictator game) is needed to get directly at these preferences. Unfortunately, a number of experiments have raised concerns regarding the robustness of dictator game results.

Oberholzer-Gee and Eichenberger (2008; OGE) test the robustness of behavior in dictator games by offering dictators the choice to play an unattractive lottery with negative expected value. Using a between groups design, there are three treatments: (i) a standard dictator game with an endowment of 7 Swiss francs (about $5), (ii) a lottery only treatment where subjects could use the 7 Swiss franc endowment to purchase a lottery ticket with negative expected

[36] For example, Glaeser et. al report that about 91 cents is returned for every dollar sent.

value[37], and (iii) an expanded dictator game in which dictators could either invest their endowment in the lottery or play the dictator game. Table 1 reports their results. In the standard dictator game median transfers are 41% of the endowment. With the lottery present the median transfer drops to zero, with the percentage of dictators keeping the entire cash endowment more than doubling and 50% of them playing the lottery rather than transfer any money. The latter contrasts strongly with the lottery only treatment where 26% of the subjects chose to play the lottery. OGE obtain similar results using University of Pennsylvania students, a $10 cash endowment, and with dictators able to invest only part of their endowment in the lottery. OGE conclude that the introduction of the lottery produces a powerful framing effect that is not explained by any extant principles. More broadly they conclude that their results imply that "…it is problematic to use the transfers observed in the context of the standard (dictator) game to make general statements about individuals' "taste for fairness."

[Insert Table 1]

List (2007) presents a simple extension of the dictator game which allows dictators to take money from receivers. For individuals who give a positive amount of money in a standard dictator game, the possibility of taking money should not affect their optimal allocation of money between themselves and the receiver.[38] In fact, giving is dramatically reduced by the possibility of taking. In the control treatment, a standard dictator game with $5 to allocate, 71% of dictators give a positive amount. When the game is extended to allow for *taking* up to $5, only 10% of dictators give a positive amount. This is consistent with a powerful framing effect.[39]

The results of OGE and List suggest that dictator game experiments can be prone to demand induced effects. Even more damaging are experiments indicating that dictator games do *not* get at outcome based preferences in isolation. Dana, Cain and Dawes (2006; DCD) report an experiment in which dictators choose a dominated alternative rather than decide how much to allocate to the second player. Their control treatment was a standard dictator game with a $10 stake. They first compare this with a treatment in which after determining how much to allocate, dictators have the option to take $9 and not play the game. If the dictator chose to make an

[37] Two different lottery treatments were employed (1) a 50-50 chance of 10 Swiss francs or 0 and (2) a 50-50 chance of 7 Swiss francs or 0. We have pooled the data and averaged across the two lotteries as the results are very similar.
[38] With well-behaved preferences over own and others' payoffs, allowing dictators to take should only affect subjects who are in a corner solution (giving zero) in the standard dictator game.
[39] See Bardsley (2008) for a dictator game that also permits taking.

allocation, the amount of money allocated along with the instructions for the game would be transferred to the second player, even if the dictator gave nothing. After making their allocation but before the money was transferred, dictators were given the opportunity to opt out of the game and receive $9 with the designated second player never even learning anything about the game.[40] Twenty eight percent (11 out of 40) dictators chose the exit option, including two who had intended to keep the $10. DCD argue that their results support the idea that some people give money in the dictator game because they are concerned with appearing to be fair recipients, a kind of audience effect.

DCD conduct a second treatment designed to rule out alternative explanations for choosing the dominated alternative: Experimenter demand effects as a result of offering the ability to opt out and exiting to escape any responsibility for the receiver. In the new treatment, in case any money allocated, only the money was transferred, without the receiver knowing where it come from (or any of the dictator's instructions). Since the receiver's knowledge is not a factor in this treatment (and dictators know this), choosing the $9 instead of making an allocation cannot be because of concerns about what receivers' think. DCD compared this "private" treatment with a replication of the exit treatment reported on above. They argue there should be significantly fewer exits in the private treatment than in the original exit treatment, since dictators know that there will be no information provided to Responders.[41] The results support their hypothesis as only 4% (1/24) dictators chose to exit in the private treatment compared to 43% (9/21) in the replication of the original exit treatment ($p < 0.01$). Some giving still occurred in the private treatment, but it was less on average than in the exit treatment. [42]

Similar concerns are raised by Dana, Weber and Kuang (2006; DWK). Consider the two step binary dictator game shown in Figure 7. In Stage 1, subjects see the box shown on top. This

[40] This was possible since the experiment was conducted as part of a larger classroom exercise with a number of other activities.

[41] To the extent that demand induced effects are present, they are present in both cases.

[42] Cherry, Frykblom, and Shogren (2002; CFS) look at a two stage game in which subjects first earned either $40 or $10 as a function of their performance on a quiz. This money was then used as the stakes to be divided in a dictator game. CFS compared the earned income condition to control sessions in which dictators were arbitrarily provided $40 or $10. They found that subjects who "earned" the money were much less generous than in the control treatment. This difference is especially striking in the low stakes treatment, since the $10 stake was earned through poor performance! CFS's results cannot be squared with an explanation of dictator game behavior that relies solely on outcome based preferences. Tadelis (2007) argues that "shame" rather than "guilt" drives the results reported in dictator games, constructing a model designed to distinguish between the two concepts and reporting an experiment based on the trust game to support his hypothesis. These are both nice papers that we encourage the interested reader to consult.

displays the dictator's payoffs for options A or B, $6 and $5 respectively, but has a question mark associated with Y's payoffs. The dictator can proceed with his choices at this point, or can click the "reveal" box, in which case they would get to see the full information regarding Y's payoffs – *either* box 1 or box 2 – before making their choice. Dictators are told that whether payoffs are determined by box 1 or 2 was decided on the basis of a coin flip *prior to the session* and that their decision to click the reveal button or not will not be revealed to Y. DWK compare this treatment with a control treatment in which dictators choose between options A and B in box 1 – with the payoffs fully revealed. In the control treatment 14 of 19 dictators (74%) chose option B, the ($5, $5) option. In the game where dictators do not know Y's payoffs, but can obtain the relevant information with a simple click of the mouse, 14 of 32 dictators (44%) chose *not* to obtain information regarding Y's payoffs. Of these, 12 of 14 (86%) chose option A with its higher own payoff. Overall, only 15 of 32 dictators (47%) chose to reveal the true state *and* chose the other regarding outcome (option B), significantly less then under the control treatment.[43] As with DCD, the results of DWK are consistent with the idea that subjects are more concerned with appearing to be fair (to themselves or others) than actually achieving a more equitable split.

[Insert Figure 7]

This idea is formalized by Andreoni and Bernheim (2008; AB). They start by noting the surprising popularity of 50-50 splits in the dictator game (about 20% of choices in Forsythe *et al*, 1994). This is difficult to rationalize with distributional models of preferences, especially given the bimodal distribution of offers: Small gifts (20% of the pie) and the 50-50 split are far more common than intermediate amounts. AB develop an elegant signaling model in which fair types use the 50-50 split to separate themselves from selfish types. To test their model they conduct a two stage dictator game. In stage one, with some probability p, nature makes the dictator's move, assigning an amount x_0 to the recipient (0 or 1 in their two treatments). Otherwise, the standard dictator game is played. Only the dictator knows the outcome of the first stage. This first stage should have no effect on play in the standard dictator game in the second stage if subjects have purely outcome based preferences, but should lead to a mass of donations at x_0 in

[43] Option A in box 2 presumably dominates all other choices so that a control treatment with box 1 and 2 both fully revealed would not serve their purposes.

the signaling model. The experimental results are consistent with the predictions of their signaling model.

It should be clear at this point that results from dictator games are sensitive to a variety of seemingly innocuous variations. This sensitivity results from several sources. First, it is well known that experimental subjects have a tendency to do what they "are supposed to do," trying to figure out what the experimenter wants and then doing it to please him or her. In other words, experiments are prone to "demand induced effects".[44] Within the standard dictator game this involves splitting the money up between themselves and an anonymous other within a social context where giving nothing is generally considered to be miserly. From this perspective, what the lottery treatment in OGE does is provide dictators with something else to do so that the desire to take an action and the desire to be generous aren't forced to be aligned. Beyond trying to please the experimenter, the results of DCD, DWK, and AB make it clear that behavior in dictator games are subject to audience effects – dictators' choices reflect a concern with how they appear to others as well as to themselves. What the DCD experiment does, for example, is to allow subjects to choose a selfish option by opting out of the game entirely, thereby not having to appear fair in the eyes the recipient or themselves. Taken together, these results indicate that dictator games cannot be treated as a Petri dish where outcome based preferences can be studied in isolation.

G. Procedural Fairness

There has been little experimental work devoted to procedural fairness in spite of the fact that data from natural experiments suggests that people are more willing to accept unfair outcomes if "fair" procedures are used to achieve these outcomes.[45] One significant exception to this is Bolton, Brandts and Ockenfels (2005; BBO). In this experiment random procedures for choosing between outcomes are introduced into the ultimatum game and what they refer to as a battle-of-the-sexes game. They note that *unbiased* random procedures capture the "level playing field" element that appears critical to many procedures that modern societies deem fair, and go on to explore the relevance of this insight to other regarding preferences.

[44] See, for example, Rosenthal and Rosnow (1969). Demand induced effects are an ever present danger in an experiment and can create a kind of Hawthorn effect when the demand effect is aligned with the treatment effect. Among other things, experimenters must be careful that their instructions and materials *not* suggest how a game ought to be played unless they have an explicit reason for doing so.

[45] See, Rotemberg, in press, for a recent survey of field data on this score.

Figure 8 shows the three games employed in experiment 1 in BBO. In the first battle-of-the sexes (BOS) game, the Proposer has only two options – a (200, 1800) allocation (option A) versus an (1800, 200) allocation (option C) where the numbers in parentheses represent the Proposers and Responders payoffs, respectively, in Pesetas. Responders have a choice of either accepting the proposed allocation (a) or rejecting it (r) and getting a (0, 0) payoff. In the ultimatum game (UG), which serves as the control treatment, the Proposer has 3 options: options A and C, the same as in the BOS game plus option B, a (1000, 1000) allocation, where rejection of any of the proposed allocation yields (0, 0). In the third game – battle-of-the-sexes with a fair procedure (BOSFP) – option B in the UG game is replaced by a new option which, if accepted, offers a 50-50 chance of a (200, 1800) allocation or an (1800, 200) allocation. Both of these outcomes are unfair but which outcome occurs is determined using a fair (random) procedure.

[Insert Figure 8]

Using the strategy method and a between groups design, the (1800, 200) option is rejected 6% of the time in the BOS game versus 41% of the time in the UG and BOSFP games. The low rejection rate in the BOS game harkens back to the idea that players do not expect Proposers to be "saints", acting against their own self-interest when they have no other choice (Bolton and Zwick, 1995), while the high rejection rates for the (1800, 200) offer in the other two cases reinforces the notion that intentions matter since in both cases Proposers could have chosen a more egalitarian option. The main point is that fair procedures are a good substitute for a fair outcome, as rejection rates for the random option in the BOSFP game were essentially the same as for the (1000, 1000) option in the UG game (1/32 versus 0 rejections in the UG case). This is an interesting line of research that deserves more attention.

H. Learning

The literature on other-regarding preferences largely takes it as given that the observed differences from classical game theory result from non-standard preferences, with the debate centered around what form these preferences take. However, models of bounded rationality and learning can provide an alternative explanation for at least some of the anomalous behavior relative standard (selfish) preferences. Although it seems unlikely that models of bounded rationality and learning can entirely explain the wide variety of other-regarding behavior observed in the laboratory, these models provide a good explanation for some phenomena which are hard to explain via non-standard preferences.

Bounded rationality and learning first entered the literature on other-regarding behavior in a pair of articles, Roth and Erev (1995; RE) and Gale, Binmore, and Samuelson (1995; GBS). Although the models used in these papers differ, the main point is roughly the same: Suppose players in the ultimatum game have completely standard selfish preferences but are adaptive learners. Rather than maximizing payoffs, players have an initial distribution over their available strategies (with the source of this initial distribution not explained). Over time, strategies that earn higher payoffs are played with greater frequency. Play in a learning model of this sort does not necessarily converge to the subgame perfect equilibrium. The logic of subgame perfection relies on players making logical inferences about play off the equilibrium path, but adaptive learning depends solely on outcomes players actually observe. If an action is taken only rarely, players never learn what payoffs would have resulted from this action and can therefore persistently play a suboptimal strategy off the (Nash) equilibrium path. In both models this is precisely the mechanism that leads to a prediction that the ultimatum game need not converge to the subgame perfect equilibrium.

Reaching the subgame perfect equilibrium under an adaptive learning model is a two step process: Responders must learn to stop rejecting low offers and then Proposers must learn that low offers will be accepted and hence are highly profitable. The timing here is tricky as Responders must stop rejecting lower offers before Proposers have stopped making them. However, the disparity in incentives between Proposers and Responders makes this highly unlikely. Rejecting a low offer costs a Responder little, but having a low offer rejected is quite costly given that roughly equal splits are almost always accepted. Proposers therefore learn to stop making low offers *faster* than Responders learn to accept them, so that play fails to converge to the subgame perfect equilibrium. This contrasts with games like the best-shot and the market game where strong out-of-equilibrium incentives push the learning process toward the subgame perfect equilibrium (which has highly asymmetric payoffs in both cases) matching the strong convergence actually observed in these games.[46]

Experimenters responded to these two learning papers by largely ignoring them, partially due to holes in the theory (particularly the source of the initial distribution of strategies) but more because of a lack of supporting experimental evidence. The theories proposed by RE and GBS

[46] The best shot game was introduced by (Harrison and Hirshleifer, 1989) and the market game is due to Prasnikar and Roth (1992). Prasnikar and Roth point out that the differing results for the ultimatum, best-shot, and market games can be attributed to differing out-of-equilibrium incentives. Roth and Erev formalize this insight.

assume Proposers and Responders learn in an identical fashion. Learning by Responders *is* predicted to be slower than Proposers' learning, but this is solely due to lower incentives to learn rather than any inherent difference between the two. Both models therefore predict that Responders' behavior should change with experience, albeit slowly. This critical prediction is difficult to test given the small changes predicted and the limited number of plays of the game in most experiments, so that power becomes a serious problem. The result is that most studies show some learning on the part of Proposers, but changes in the behavior of Responders is generally too small to be statistically significant (see SR, 1998, for example).

However, two papers that are specifically designed to provide more powerful tests of Responder learning find evidence for it.[47] Cherry and List (2000; CL) run a variant of the SR's high-stakes experiment where Proposers earn the right to propose rather than having it determined exogenously. This results in substantially more low offers than is typical – 28% of offers are less than a quarter of the pie. With an enlarged sample of low offers, CL find that (controlling for the size of offers) rejection rates fall with experience. However, this result does not represent unqualified support for the learning models as both predict that learning should slow down over time. This is an explicit feature of the RE model and also holds for GBS due to the changing distribution of offers. Most of the decline observed in CL occurs in the last few periods, which is more consistent with a reputation model in the spirit of Healy (2007) than an adaptive learning model.

Cooper, Feltovich, Roth, and Zwick (2003; CFRZ) manipulate the experience received by Responders by doubling the number of Proposers relative to Responders.[48] Since Responders are playing twice as often as Proposers, the relative speed of learning argument that drives the predictions of RE and GBS should have less force. Indeed, CFRZ find that rejection rates are lower in the 2 x 1 treatment than in controls which have an even number of Proposers and

[47] Two other papers that deserve note are Winter and Zamir (1997; WZ) and Armantier (2006). In WZ Responders unknowingly play against a mixture of real and computer opponents who are programmed to make low offers. They find that Responders in this environment learn to accept all offers more frequently than Responders who face a mixture of computer and human opponents who make high (almost even) offers. The original version of this paper generated some controversy due to subjects not knowing they were playing against a mix of computers and humans. Subsequent versions of the paper included treatments in which subjects were told the percentage of offers that were generated by the computers, with essentially no difference in outcomes between the two cases. Armantier primarily focuses on the interaction between the initial distribution of wealth and fairness, but also fits a reinforcement learning model to his data. This fitting exercise provides evidence for learning by Responders.

[48] The experiments are designed so subjects have no way of distinguishing whether they are in a standard session where the number of Proposers and Responders are even or in a treatment session where the number of Proposers is doubled.

Responders. They also find that a history of receiving low offers makes subjects more likely to accept low offers, and that the treatment effect does not widen over time, consistent with learning slowing over time. Although these results are consistent with adaptive learning, the evidence is indirect as the treatment effect cannot be observed with the naked eye and the magnitude of the estimated effects is moderate.[49]

To directly test the prediction that rejection rates fall with experience in the ultimatum game, Cooper and Dutcher (2009; CD) have done a meta-analysis pooling data gathered in six different experiments.[50] Their criteria for choosing studies was as follows: data had to be from "standard" ultimatum games (i.e. played with direct response rather than the strategy method, random re-matching between rounds, random selection into roles, endowments are provided exogenously, Proposers and Responders play with equal frequency) and subjects had to play at least ten rounds. The resulting data set has observations from 365 Responders.

The main result of this meta-study can be seen in Figure 9. The x-axis gives the proportion of the pie offered to the Responder and the bars show acceptance rates. The labels at the top of the bars show the frequency of each offer category. Since only three of the six datasets have more than ten rounds, this figure compares data from Rounds 1 – 5 with data from Rounds 6 – 10.[51] The overall pattern is clear. The acceptance rate rises with experience for relatively large offers, but falls for small offers (20% of the pie or less).

[Insert Figure 9]

The magnitude of these changes is small, but the advantage of doing a meta-study is that the large data set provides the necessary power to detect small changes. CD run appropriate regressions, controlling for session and individual effects, and confirm that the changes in acceptance rates with experience, both the increase for relatively large offers and the decrease for small offers, are significant at the 1% level. The regressions also indicate that there is no learning beyond the first ten periods, consistent with the learning models' prediction that learning should slow down with experience.

CD establishes beyond a reasonable doubt that Responder behavior in the ultimatum game changes with experience. The effect is small, but that is predicted: Responders learn the

[49] Other papers looking at learning in ultimatum games using clever experimental designs to deal with the power issue are Abbink, Bolton, Sadrieh, and Tang (2001) and Andreoni and Blanchard (2006).
[50] Data were provided from Roth et al. (1991), SR (1998), Duffy and Feltovich (1999), Anderson, Rodgers, and Rodriguez (2000), CFRZ (2003), and Andreoni, Castillo, and Petrie (2007).
[51] Expanding the dataset to include all observations has no effect on the conclusions.

same way as Proposers, but are much slower to learn as the incentives to learn are lower. What is *not* consistent with the learning models discussed so far is the reduced acceptance rates for the very lowest offers shown in Figure 9. This suggests that the adjustments observed in behavior over time reflect something beyond simply learning to follow the money-maximizing strategy of accepting all offers. CD note that this pattern is consistent with learning in a Charness-Rabin framework where Responders' perceptions of kindness depends on their belief about the distribution of offers. As Responders learn about the distribution of offers, their beliefs and hence their actions should change. CD present evidence based on individual subject data that is consistent with such a model.[52]

An underappreciated source of evidence against *pure* adaptive learning models (i.e. models that only allow for standard selfish preferences) is results from gift exchange experiments. In these experiments, discussed in detail in section III, employees consistently provide higher effort in response to higher wages. This behavior is irrational in the same way that rejections are irrational in the ultimatum game – giving positive effort is not payoff maximizing and cannot be attributed to the effects of strategic uncertainty. There is however, a critical difference. Responders in ultimatum games get few opportunities to respond to low offers. In contrast, subjects in gift exchange experiments are frequently offered wages well above the reservation level and get many chances to learn that they need not respond with high effort. Thus, the argument for slow learning by Responders that applies to ultimatum games should not work in gift exchange games. Rather Responders should steadily learn to provide less effort. However, effort only tails off (if at all) in the last few rounds of a gift exchange experiment, a result that is more consistent with a reputation model (Healy, 2007) than adaptive learning. Thus, pure adaptive learning models fail a critical test of their ability to predict across related games.

While observed changes in behavior are small for Responders in the ultimatum game (and apparently non-existent for employees in gift exchange games), changes in behavior can be quite large in other, related, situations. For example, Cooper and Stockman (2002; CS) study a three player sequential step-level public goods game. Players take turns deciding to contribute or not contribute to a public good. The good is provided if two or more players decide to

[52] There is a strong negative relationship between the lagged offer and the likelihood that the current offer will be accepted. This is consistent with subjects lowering their expectations about the distribution of offers after receiving a low offer.

contribute. Critically, costs of contribution are sunk and rising in the order of play. For all treatments the value of the public good is 18 tokens if provided. The cost of contributing for each of the three players is 3/6/9, 1/3/9, and 1/3/16, respectively, across their three treatments. CS focus on the behavior of "critical" third players, players whose decisions determine whether or not the public good will be provided. Similar to Responders in the ultimatum game, critical third players face no strategic uncertainty. As in the ultimatum game, there is a tension between payoff maximization and fairness – critical third players always make the most by contributing but always make less than the other two players if they contribute. For all treatments CS find that contribution rates change significantly for critical third players with experience. The surprise is that with the most uneven payoffs, as in the 1/3/16 treatment, contributions rates for critical third players *fall* sharply with experience (see Figure 10). This indicates that there is a significant dynamic to be explained and that the explanation cannot rely purely on adaptive learning, but may instead involve a combination of adaptive learning and other regarding preferences. CS show that a hybrid model combining adaptive learning and other-regarding preference can rationalize their result. The dynamics observed in CD, growing acceptance of moderately uneven offers along with growing rejections of the most uneven offers, are also consistent with this hybrid model.

[Insert Figure 10]

Summary: Pure adaptive learning models can predict the main features of behavior in the ultimatum game (as well as more subtle ones such as the differing speeds of learning for Proposers and Responders). However, these models alone cannot provide a complete explanation of other regarding behavior as they cannot, by design, explain initial behavior. For the ultimatum game, knowing that the Responders' behavior changes slightly is less important than the fact that sizable positive offers are frequently rejected even among experienced players. Nonetheless, learning models (adaptive and otherwise) remain an important topic in the study of other-regarding behavior for two reasons. Even though the observed changes in behavior can be quite small, these changes reinforce our impression that behavior observed in laboratory studies cannot be interpreted as directly revealing stable, complete preferences. Second, since the observed changes in behavior vary both in magnitude (compare the dynamics in the ultimatum game with those observed by CS) and direction (toward or away from the money maximizing choice), they point to the need to develop a model which can explain this variation without

relying on the resolution of strategic uncertainty or reputation effects. Promising candidates for such a model include hybrid models like the one proposed by CS and models of social norm formation where changes of behavior are driven by changes in beliefs about what constitutes misbehavior, but there is a lack of evidence distinguishing between these models.

I. Other-Regarding Behavior and Utility Maximization

Much of the existing literature on other-regarding behavior revolves around attempts to identify the preferences underlying seemingly anomalous behavior. As such, it can be characterized as neo-classical economics flavored with a dash of psychology. Subjects are presumed to be maximizing a stable utility function, with the theory departing from standard microeconomics only through the arguments in the utility function. Even theories which have roots in psychological game theory (e.g., CR, 2002) rely on subjects maximizing utility subject to stable preferences. The work described in this section directly addresses the question of whether or not other-regarding behavior is consistent with rational choice theory as understood by economists.

Andreoni and Miller (2002; AM) address this question in the most direct possible fashion and provide a strong affirmative answer. Subjects made decisions in a series of modified dictator games. Both the available budget and the relative price of giving were varied across games. In other words, subjects were asked to choose between payoffs for themselves and payoffs for another anonymous subject under a variety of budget constraints. Rather than testing any particular theory of other-regarding preferences, AM focus on whether choices are consistent with the generalized axiom of revealed preference.[53] They find that a remarkable 90% of the subjects have no violations of GARP (with at least eight choices per subject), implying that most subjects' choices are consistent with maximization of a quasi-concave utility function. Even though 23% of the subjects never gave away any money, most of the cases where subjects were giving money away are compatible with a rational choice model in terms of satisfying GARP. AM also note that there is a great deal of heterogeneity among subjects – beyond the large number of purely selfish subjects (47%) who kept all the money to themselves, there were sizable numbers of subjects who always split the money equally between themselves and the

[53] A is directly revealed preferred to B if A is chosen when B was an available choice. A is indirectly revealed preferred to B if there is a chain of directly revealed preference running from A to B (e.g. A is directly revealed preferred to C is directly revealed preferred to B). GARP states that if A is indirectly revealed preferred to B then B cannot be strictly directly revealed preferred to A.

other player (30%) or gave all the money to the player with the highest payoff (22%).[54] This heterogeneity needs to be taken into account when looking at other regarding behavior. Finally, AM present evidence that behavior reported in other other-regarding experiments could have been (approximately) generated by the distribution of preferences they report. They argue that this is evidence that preferences are robust over a variety of settings.

Fisman, Kariv, and Markovits (2007; FKM) provide a more powerful test of GARP as subjects are asked to make fifty decisions rather than the eight used in most of AM's sessions. Looking at decisions in two person modified dictator games, the proportion of subjects whose decisions are completely consistent with GARP falls to 11%. This decline relative to AM's data is to be expected given the substantial increase in the number of decisions. However, FKM conclude that violations of rationality are generally small as 86% of subjects have CCEI scores, which measure how much a subject's budget constraint would need to be perturbed to make their choices consistent with GARP (Afriat, 1972), of .8 or greater.[55] FKM also expand AM's analysis of individual utility functions by studying three person dictator games, so that they can address broader classes of other-regarding preferences. While they continue to find evidence in favor of social welfare preferences, their primary conclusion is that preferences are quite heterogeneous.[56]

The results of AM and FKM make a good case for other-regarding choices being consistent with maximization of a well-behaved utility function. Both show that preferences are reasonably consistent with standard theory in stable environments in that other regarding behavior is price sensitive in the usual ways. Further, there is a good deal of heterogeneity in the preferences with large numbers of subjects having standard selfish preferences and others having other regarding preferences. Both studies rely on environments that, other than changing prices and budgets, are stable. This misses some of the key problems already identified in terms of standard utility functions: Altering seemingly irrelevant features of the decision making environment which often change choices (see III.F above) or when the situation is sufficiently non-trivial that learning is involved (see CS, 2002, in III.H above).

[54]These pure cases account for a total of 43% of their subjects, *always* choosing in the ways specified. The remaining 57% of subjects are classified as "weak" types by determining the distance of their choices from those of each pure type.
[55] The scale runs from 0 to 1 with numbers closer to 1 indicating choices that are more consistent with GARP.
[56] In the three player case the proportion of subjects whose behavior is completely consistent with GARP rises to 25%. Only 12% have CCEI scores below .8.

IV. Gift Exchange Experiments

A. An Initial Series of Experiments

In a remarkable series of experiments, Ernst Fehr and his colleagues explored behavior in the *gift exchange game*. Although the concept of gift exchange applies to a variety of economic settings, for the sake of clarity we use a labor market framework to characterize the structure of the game. The typical gift exchange game is a two stage game. In stage 1 employers' make costly wage offers to potential employees. In stage 2 employees decide to accept or reject the proposed offer and then provide a costly "effort level" to employers, with more effort being more costly. The higher the effort level provided the greater the employer's profits are. The game is usually repeated over a finite number of trials, with the number of trials announced in advance. Matching of firms and workers are anonymous so that there is no opportunity for workers to develop individual reputations, or for other repeated game effects to occur.

Firms and workers are provided with payoff functions of the following sort:

$$\Pi_M = (v - w) * e \quad (1)$$

$$\Pi_E = w - c - m(e) \quad (2)$$

where M represents the manager, E the employee, e denotes the employee's effort, w is the wage, and m(e) is the cost of effort. In the original Fehr, Kirchsteiger, and Riedel (1993; FKR) paper v was set at 126, c was set at 26, and m(e) was determined according to the following table of values.

Effort	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Cost	0	1	2	4	6	8	10	12	15	18

A variety of different matching formats have been used for firms and workers in this game including one sided posted (or oral) offer auctions, continuous double auctions, and one-to-one matching with different partners. In the auction markets there are typically excess numbers of workers so prices above the minimum wage cannot be explained by workers exercising market power.

Since there are a finite number of plays of the game with a known end point, standard unraveling arguments predict a minimum wage offer in all periods accompanied by minimum effort levels. Yet there is typically a clear positive relationship between wages and effort levels

(see Figure 11), resulting in a Pareto improving outcome with higher earnings for both firms and workers than absent gift exchange. These results do not appear to depend on the fine points of the market institutions (e. g., posted offer vs. double oral auction or one-to-one matching).

[Insert Figure 11]

The results are consistent with Akerlof's (1982) model of gift exchange, in which employers receive higher productivity from employees by paying them above market wages. But this result emerges in a much starker environment than the one Akerlof describes, as workers are isolated from each other so that social norms regarding appropriate effort levels cannot emerge, and there is no potential to fire employees whose effort levels fall below the firm's expectations. Since these mechanisms cannot cause the positive relationship between wages and effort, Fehr and his colleagues explain their results via positive reciprocity between firms and workers, consistent with the positive reciprocity observed in the trust game.

Fehr and his colleagues have applied the gift exchange game to a variety of interesting issues in labor economics. Fehr and Falk (1999; FF) showed that wages fail to converge to the competitive equilibrium level in a continuous double auction labor market in spite of more or less continuous efforts to undercut wages by unemployed workers. Firms apparently refuse the lower wage offers since they tend to be accompanied by lower effort levels, to the point that it is not profit maximizing to hire such workers.

The preceding raises a puzzling feature of gift-exchange experiments. Employers who offer wages above the minimum earn positive profits, in contrast with the trust game where sending money is a break-even proposition at best (e.g. BDM, 1995). This is an important point since Fehr and company argue that the profitability of wages above the minimum is a major reason for their persistence. In contrast, we suspect that this difference between the two classes of games can be explained by differences in the costs of reciprocal behavior. In the trust game, the receiver must spend a dollar to give a dollar to the sender. Compare this with FF's (1999) gift exchange game when a wage of 60 (roughly the mean wage) is chosen. In this case, at low effort levels, an employee giving up a dollar in own payoffs raises the employer's payoff by up to six dollars! Consistent with the results of AM (2002), there is greater reciprocity when the cost of kindness is cheaper.

Remark: The responsiveness of reciprocal behavior to costs has implications for the growing number of experiments investigating gift exchange in field settings (see section IV.C below). In particular it creates a major problem for exploring gift exchange in field settings where neither

the cost of effort, nor the ratio of cost to benefits generated by the effort, can be readily measured. Without this it cannot be determined if a failure to observe gift exchange in a particular environment reflects employees have selfish preferences or that the cost of reciprocating is too high relative to the benefits, or that employees have found a less costly way of rewarding employer's kind behavior.

The existence of a successful gift exchange paradigm in the laboratory makes possible controlled investigation of a number of interesting issues in labor economics. We briefly discuss two applications below.

One branch of the gift exchange literature deals with the implications of voluntary gift exchange for implicit versus explicit labor contracts. This is interesting since the implicit contract literature points out that it is essentially impossible to write fully enforceable complete contracts. Further, since there is some suggestion in the literature that the introduction of explicit incentives for a given behavior may drive down intrinsic incentives (Deci, 1999), what these experiments look at is whether or not it is desirable to have as complete contracts as possible. Fehr and Gachter (2002; FG) report one such experiment. There are three treatments: (1) Firms specify a wage along with a desired effort level and a fine if that effort level is not realized, with an exogenously determined one-third probability of identifying shirking workers, (2) Firms specify a bonus along with a wage and a desired effort level, with an exogenously determined one-third probability of the bonus not being paid in case the worker was identified as shirking, and (3) a control treatment in which the firm simply specifies a wage along with a desired effort level.

Their results are reported in Figure 12 with effort levels reported on the vertical axes and wages in excess of the competitive equilibrium level reported on the horizontal axes. The fine treatment generated consistently lower effort levels at each wage rate than with no explicit incentives. The bonus treatment did better than the fine treatment, but the standard gift exchange treatment without any explicit fines or bonuses did even better. Although explicit incentives resulted in lower effort levels in this experiment and lower social surplus, employers' profits were higher in the fine treatment as much lower wages were offered while extracting higher effort levels out of workers in response to the probability of being hit with a hefty fine.[57]

[57] A number of other papers explored this and related issues: Fehr, Gachter, and Kirchsteiger (1997), Brown, Falk and Fehr (2004), Falk and Kosfel (2006), Fehr, Hart and Zender (2008), and Ploner and Ziegelmeyer (2008). This line of research is closely related to the general issue of whether explicit incentives for a given behavior drives down extrinsic motivation.

[Insert Figure 12]

Surveys conducted to investigate reasons for wage rigidity in the face of rising unemployment commonly report that employers fear adjusting wages downward will result in negative effects on employee effort, along with adverse selection with respect to quits (Campbell and Kamlani, 1997, Bewley, 1998). The validity of this fear isn't obvious. Models of outcome based preferences predict that employers faced with bad business conditions (i.e. low revenues) should be able lower wages without reducing employee effort, and even models that allow for reciprocity admit the possibility that a wage cut won't harm employee effort under these circumstances.[58] To sort this out, Hannan (2005) presents an experiment studying the interaction between firm profitability and employees' reactions to wage changes. Her experimental design employed payoff functions similar to (1) and (2) above, with the typical two stage process where firms first set wages and employees respond with effort. However, she added a third stage: After firms set wages and workers responded with their effort level, but before these effort levels were reported back to employers, there was an exogenous shock to the firm's profit – a random draw with a one-third probability of a positive, negative, or zero shock.[59] If there was no shock, the wage and effort levels agreed to originally were binding and the round ends. If there was a positive or negative income shock it was publicly announced, and firms and workers had the ability to adjust their wages and corresponding effort levels.

Figure 13 reports her main results, with wage changes reported on the horizontal axis and the change in mean effort levels on the vertical axis. The data support the fear that adjusting wages downward, even after a negative income shock, will result in lower effort levels and be unprofitable for employers as: (1) When wages are decreased, workers tend to decrease effort, regardless of whether or not the income shock is positive or negative. (2) The magnitude of the punishment or reward, in terms of effort, is directly related to the magnitude of the wage change, regardless of whether or not there was a positive or negative profit shock. (3) The magnitude of the negative response to wage decreases is twice that of the positive response to wage increases so that workers punish firms more for decreasing wages than they reward them for increasing

[58] In any such model, effort depends on whether the wage is considered kind or unkind. If the employer's profitability falls, lower wages could be considered less unkind than previously.

[59] Income shocks were set at two different levels depending on the experimental session, and represented 50% or 100% of the lowest possible wage, and 8.3% or 16.7% of the highest possible wage.

wages. (4) Firm profits were significantly lower if they reduced wages following a negative profit shock than if they had held wages constant.

[Insert Figure 13]

B. Why Does Gift Exchange Occur?

Fehr and his colleagues have used the results of the gift exchange game to argue for the importance of positive reciprocity and the desire to avoid social disapproval in economic interactions (Fehr and Falk, 2002). Bolton and Ockenfels (2000) analysis of the gift exchange game postulates a heterogeneous population made up of egoists that maximize pecuniary payoffs and other regarding income types who reciprocate with greater effort in response to higher wages provided they are able to get at least half the pie. An increase in wages increases average effort as the other regarding types respond with higher effort and the egoists' effort level remains constant, so that average effort increases. And indeed, in most gift exchange experiments there are a minority of subjects who more or less continuously provide minimal effort, regardless of the wage rate. Further, the BO model predicts that higher wages will not always be met by higher profits for firms, as other regarding workers generally insist on getting at least half the efficiency gains from reciprocating with higher effort, which is not always possible given the (typically) increasing cost of higher effort for workers. This too is found in the data (BO, 2000).

Other studies have failed to find the same high levels of gift exchange reported by Fehr and his colleagues. For example, Hannan, Kagel and Moser (2002; HKM) compare behavior in a gift exchange game using undergraduates and MBAs, as well as worker responses to high versus low productivity firms (where high productivity firms find it less costly to provide higher wages than do low productivity firms).[60] Wage offers were tagged with the firm's productivity level in a posted offer labor market. They find no difference in worker response to comparable wage offers from high versus low productivity firms, as well as a marked difference in effort levels between undergraduates and MBAs. Undergraduates provided substantially lower effort levels than both the MBAs and the effort levels reported in FKR (1993), particularly at higher wage rates (see Figure 14).[61] They conjecture that the lack of responsiveness to differences in firm's productivity levels (which was true for MBAs as well as for undergraduates) resulted from a lack

[60] $\Pi_M = (v\text{-}w)\, e$ where $v = 90$ for low productivity firms and 120 for high productivity firms, so that for any given wage-effort level payoffs to high productivity firms were higher than for low productivity firms.

[61] Note, HKM do not assert that undergraduates do not provide statistically significant higher effort levels at higher wages. Just that they are considerably lower than found with MBAs and in FKR (1993).

of saliency, as the relationship between firm profits and productivity is an indirect one. As for the difference between MBAs and undergraduates, they note that MBAs have greater experience in jobs where gift exchange plays an important role so were more able to relate their past experience to the labor market context under which the experiment was conducted. In contrast, most undergraduate work in the United States is associated with minimum wage jobs where there is no, or minimal, gift exchange. This interpretation is consistent with the Akerlof (1982) model of gift exchange which assumes that higher wages result in higher effort levels out of social norms and conventions in the workplace. Only in this case, it is conventions and norms from the workplace that carry over into the lab.

[Insert Figure 14]

Healy (2007) formalizes the role of reputation in fostering gift exchange. On first blush reputation does not seem to be a likely explanation for the positive relationship between wages and effort in gift exchange experiments. Matchings in these experiments are typically anonymous, meaning employers are unable to track an employee's behavior over multiple rounds, and the pool of employees is sufficiently large that no one employee greatly impacts the reputation of the pool as a whole. Healy's innovation is to note that stereotyping greatly enhances the importance of reputation building. Stereotyping refers to the (possibly irrational) attribution of characteristics of a group to individuals within the group even if the group members are known to be heterogeneous. Within the gift exchange framework, suppose there are two types of employees, strictly selfish types (egoists) and reciprocal types who will reciprocate with higher effort in response to higher wages (reciprocators). If the probability of reciprocal types is known and types are independent across individuals, observing reciprocal behavior from one employee reveals nothing about the extent to which other employees are reciprocal. With stereotyping, types are believed to be positively correlated across employees. Observing reciprocal behavior increases the perceived likelihood of reciprocal behavior from all employees. If the perceived positive correlation of types across employees is sufficiently strong, it becomes worthwhile for even egoists to exhibit reciprocal behavior until the final period of the experiment so as to maintain the group's reputation for reciprocity and the resulting high wage offers. Reputation building can therefore explain a significant portion of the positive relationship between wages and effort in gift exchange experiments.

While the primary contribution of Healy is theoretical, he also presents a series of experiments that test predictions of his reputation model. The model predicts that effort should collapse in the final round of play, since there is no benefit to maintaining a reputation for egoists, and that if payoffs are manipulated to require a higher probability of reciprocal types to maintain a reputation equilibrium it is possible that the positive relationship between wages and effort should collapse. The experimental results are largely consistent with these predictions.[62]

Healy's work leaves a number of open issues. For example, it does not provide an explanation for the sizable population of consistently selfish employees observed in most gift exchange experiments (in a reputation model, these individuals should be imitating the reciprocating types). It would also be nice to have direct evidence of the stereotyping on which the model relies so heavily. Nonetheless, Healy makes it clear that positive reciprocity is *not* the only plausible explanation for positive correlation between effort and wages in gift exchange games.

C. Laboratory vs. Field Settings and Real Effort

One relevant question is whether or not gift exchange carries over to environments where workers have to respond with real effort as opposed to the higher monetary costs associated with greater "effort" in the typical experiment. An initial answer in the affirmative was provided by Gneezy (2004) who used solving mazes as his real effort task, employing mazes with different levels of difficulty (as measured by average time to solve a maze) and with different returns to "firms" for each maze solved. In a single period game he found that in all treatments the higher the wage the higher the number of mazes solved, consistent with the presence of positive reciprocity. However, higher wages did not always result in higher profits.

In a more recent experiment Gneezy and List (2006; GL) look at two tasks – computerizing library holdings over a 6 hour period and a door-to-door fundraising effort over a single weekend day. The gift exchange treatment was operationalized by advertising a given wage rate and then when subjects showed up, paying a higher than advertised wage for one of the two groups; e.g., an advertised wage of $12 per hour for the library task, with half the subjects given the "surprise" wage of $20 an hour upon showing up. Procedures were designed to insure that subjects in the two treatments were not aware of the difference in wage rates.

[62] In some sessions, effort begins to collapse prior to the final round. This may reflect "trembling" on the part of players or could reflect a mixed strategy equilibrium – while Healy focuses on a tractable pure strategy equilibrium, the game supports mixed strategy equilibria as well, with the probability of cooperation decreasing over time.

Figure 15 reports the results from the library task split into 90 minute intervals. In the first 90 minutes the average number of books logged into the computer is significantly higher for the high wage group but this trails off over time, with the averages the same over the last 3 hour period. Similar results were found in the fundraising task in terms of the amount of money raised, with significantly larger amounts raised before lunch in the high wage treatment compared to the advertised treatment. But this difference was small and not statically significant after lunch. GL interpret these results in terms of the psychology literature on reference point effects, arguing that after a while workers reference point shifts so that the new higher wage serves as the fair wage reference point, with a resulting drop in effort. These results, particularly given their interpretation, set off a fire storm since they seemed to be at odds with the large body of laboratory research on gift exchange. However, there are several alternative explanations for GL's results that would be consistent with the laboratory research: (1) the higher wage workers became fatigued from working harder and/or (2) the higher wage workers provided the gift level they thought appropriate to the higher than advertised wage in the first half of the day and slacked off after that. All three interpretations (and no doubt others) are consistent with the data.

[Insert Figure 15]

More importantly, other field experiments have reached different conclusions. Kube, Marechal, and Puppe (2006; KMP) look at gift exchange in a library cataloguing task, focusing on both positive and negative reciprocity. Students were hired for a six hour shift at cataloguing books, with the recruitment e-mail announcing a *presumptive* salary of 15 Euros per hour. Upon arrival one-third of the subjects were told the wage would be 20 Euros per hour (the "Kind" treatment), with another third told that the wage was actually 10 Euros per hour (the "Unkind" treatment), with the last group getting the 15 Euro per hour wage (the "Neutral" treatment). Subjects in each group did not know what others were paid.

Figure 16 reports the average number of books catalogued for each group by 90 minute blocks. The number of books catalogued rises over time for all three treatments. The Unkind treatment starts out at a much lower rate of cataloguing than the Neutral treatment and remains below it throughout, with the differences statistically significant in each 90 minute interval. The Kind treatment starts at the same rate as the Neutral treatment but catalogues at higher rates for each 15 minute interval after that, although with the exception of the middle interval (between

minutes 181 and 270) these difference are not statistically significant at conventional levels. KMP conclude, as many do, that negative reciprocity is a stronger force than positive reciprocity (see, for example, Rotemberg, 2006, who summarizes field data to this effect). More to the point, unlike the pattern reported in GL, positive reciprocity is growing over time. This study has obvious weaknesses as the difference between the Kind and Neutral treatments are not statistically significant at conventional levels, although the lack of statistical significance could be attributed to the small sample sizes (9 and 10 subjects in each treatment). Be this as it may, the qualitative results of KPL stand in strong contrast to those of GL as negative reciprocity is strong throughout and positive reciprocity *grows* over time rather than tailing off.

Gneezy and List themselves are co-authors on another study with qualitative results opposite to those GL report. Al-Ubaybli, Andersen, Gneezy, and List (2007) report a field experiment with temporary workers in which effort levels are growing over time in both the baseline treatment and the positive gift exchange treatment, with the differences in favor of the gift exchange treatment increasing over time, to the point that they appear to be statistically significant at the end of their two day trial.[63]

Bellemare and Shearer (2007) find that an unexpected one-time bonus for workers in a tree planting firm significantly increases the number of trees planted, with the response from workers increasing with their tenure in the firm. One strength, as well as weakness, of this study is the long term relationship the workers had with the firm. From a labor economics point of view, these sorts of long term employer-employee relationships are what Akerlof had in mind when he was originally writing about gift-exchange. It is therefore especially valuable to see that strong positive reciprocity is present in such an environment. However, from an experimental point of view, the long term interaction of firms and workers makes it difficult to determine whether gift exchange is occurring due to subject preferences, game theoretic concerns (e. g., maintaining reputations, avoiding punishment in a supergame), or some combination thereof. This illustrates a general problem with the field studies of gift exchange, as workers may think

[63] Unfortunately, the statistical specification omits any interaction terms for the time trend variable with the treatments which would reveal whether or not effort for the gift exchange treatment was growing faster over time than the controls.

they are playing a different game than the experimenter has in mind – even temporary workers may believe there is the possibility of a long term relationship.[64]

To summarize, the evidence from field experiments on gift exchange is mixed. This sub-literature illustrates both the strengths and weaknesses of field experiments. The subjects in these studies do not know they are in an experiment, the tasks they are being asked to perform parallel those they would normally perform, and, in some cases, the experiment takes place in the context of a longer term relationship. As such, these experiments should be less prone to demand induced effects than laboratory experiments and also should have a closer relationship to the field setting that authors like Akerlof had in mind. However, the cost of this verisimilitude is high. There is a tremendous loss of control in these experiments, as we neither know the cost of effort, the perceived benefits of effort to the employer, nor the game that the employees think they are playing. Measurement is a problem in many of these studies as workers in field settings can respond to incentives along multiple dimensions, so that the experimenter may miss important elements of employees' responses to a gift. Also one must take account of the level of baseline wages relative to market wages for comparable work as higher than normal baseline wages may already elicit a strong gift response. As Cohn et al. (2007) note, this may impose a ceiling effect resulting in a downward bias in the response to the gift wage treatment.[65] Finally, while it is clear that subjects in a laboratory choosing numbers to represent effort are performing a substantially different task than a worker planting trees in British Columbia, it is not so clear that one of these cases is closer than the other to the situation of stock-brokers working in a Boston office. All settings have specific elements which may affect behavior.

D. Summary

The experimental literature on gift-exchange has been highly influential, and deservedly so. Even if gift exchange does not always occur in either laboratory or field studies, it occurs often enough and is strong enough to be an important phenomenon. It remains an important question to determine why gift exchange manifests itself in some settings and not in others. We conjecture that the answer to this question will largely be economic in nature, relating to the costs of reciprocity and the game that subjects perceive they are playing. With respect to the latter, it would be interesting to know if the experience effects reported in HKM are due to

[64] See Cohn, Fehr, and Goette (2007) for an experiment that measures players' fairness perceptions in a field experiment.

[65] Unfortunately, the relationship of baseline wages to market wages is not always reported.

changes in preferences that MBAs undergo as a result of prior work experience compared to undergraduates, or to perceived differences in the game being played. Finally, an important open issue is how gift exchange will work when multiple avenues of reciprocity are open. Will employees focus on the cheapest method of reciprocating, or will they also consider the benefits to employers in determining how to reward the gift of above market wages?

Conclusions

The experimental literature on other-regarding behavior has been extraordinarily rich and abundant over the past ten years, and will no doubt continue to be going forward. The present survey is selective as there are many fine papers that we could have, at the risk of completely overwhelming readers, reviewed. Rather our goal has been to cover the range of research, to identify some of the highlights as well as some of the deficiencies in the existing literature, and to make some heretofore overlooked connections between different branches of the literature.

There have been some clear successes in the past ten years or so: there now exist well developed theories of outcome based preferences and reciprocity, an increasingly detailed picture is developing of when other-regarding behavior is and is not likely to occur, and, particularly through the literature on gift-exchange, it is becoming increasingly clear why the laboratory studies of other-regarding behavior are important to mainstream economists. There are also a number of issues that remain to be resolved in this literature: It is clear that none of the existing models fully capture the determinants of other-regarding behavior, and those that attempt to threaten to lose tractability. There is a tremendous amount of procedural variation in studies that aim to look at similar phenomena. In particular there is an over reliance on the strategy method with their one-shot "what if" approach as opposed to behavior resulting from experienced play. It is clear that other-regarding behavior is sensitive to the context in which it is studied which makes it difficult to determine how results will generalize from one setting to another.

One way we could have concluded this chapter is by giving a laundry list of important questions that remain open. We're not going to do that since the exercise would be a bit like a broker giving stock tips – if the research ideas were really good, why would we be sharing them? We could take one final stab at what it all means, but at this point in time it's not entirely clear as the literature on other-regarding behavior has a long way to go yet.

Instead, we thought it would be fun to take the unusual step of each of us giving a list of eight papers from this literature that every experimental economist should read, even if they read

nothing else. Given the many papers that have been written on the topic it can be hard to see the forest for the trees, and the literature on other-regarding behavior and preferences has a lot of trees. So here are our admittedly idiosyncratic takes on what you should read and why. We are limiting ourselves to papers not covered in the earlier *Handbook of Experimental Economics* (1995) and, just to avoid an obvious incentive problem, neither of us will choose papers on which we were co-authors.

Cooper's Elite Eight

1) Fehr and Schmidt (1999)/Bolton and Ockenfels (2000): The two most important models of outcome based preferences. These tie together much of the preceding literature and place a literature that often felt like bad pop psychology on a firm foundation of economic theory. Which paper should take pride of place? Flip a coin.
2) Blount (1995): An elegant experiment that makes it completely obvious why outcome based preferences are not enough.
3) Charness and Rabin (2002): The most influential model of other-regarding behavior that incorporates psychological game theory, plus an important generalization of existing models of outcome based preferences. Read the working paper version – it's better than the published version.
4) Fehr, Kirchsteiger, and Riedl (1993): If there is any one reason why economists who are not experimenters should care about other-regarding behavior, the literature on gift-exchange is it. This is the paper that started this strand of research in experimental economics.
5) Andreoni and Miller (2002): This paper shows that the standard economic theory we all learned in our first semester of graduate school still remains relevant in the brave new world of other-regarding preferences.
6) Kagel and Wolfe (2001): This illustrates one of the biggest problems in the literature on other regarding behavior and preferences – who is the "other"?
7) Dana, Weber, and Kuang (2007): This paper is one of a group of papers that elegantly and persuasively establishes the idea that other-regarding behavior, particularly in dictator game, can be quite sensitive to how subjects think their actions will be perceived, both by themselves and others. These papers vividly illustrate the delicacy of dictator

game results, hopefully putting an end to the misuse of that particular instrument. John picks the other papers in this group below, but I like this one best. Read them all – taken together they provide a damning critique of over-reliance on dictator game experiments.

8) Xiao and Hauser (2006): It has communication, so how could it possibly not be interesting? More to the point, a weakness of the literature on other-regarding behavior is that individuals are invariably given only a single method of rewarding and punishing others. This paper illustrates how much outcomes might change when a richer and more realistic set of options is available.

Kagel's Elite Eight:

1.) Fehr and Schmidt (1999)/Bolton and Ockenfels (2000): It was clear at the time that both these papers were written that they had to be "wrong", but as one of my old teachers used to say "wrong in the right way." Both papers pulled together a large number of experimental studies into a coherent framework without having to ignore too inconsistencies. They motivated a host of new experiments which have enriched our understanding of behavior.

2.) Blount (1995): This was the first paper to make it clear that intentions mattered.

3.) Charness and Rabin (2002): This has been one of the more influential psychological game theoretic models of other regarding behavior. It also introduced the notion of maximin preferences and taste for efficiency which set off a number of new experiments exploring these issues.

4.) Oberholzer-Gee and Eichenberger (2007): This paper has actually been around in one form or another for quite some time. It was one of the first papers demonstrating the instability of dictator game choices as opposed to what the game was originally intended to do – provide a clean test of the hypothesis that near equal splits in the ultimatum game were not a result of Proposers trying to be fair to Responders, but rather a strategic response to anticipated rejections of low offers. Runner up here would be Dana, Cain and Dawes (2006).

5.) Prasnikar and Roth (1992): I've broken the rules with this one as it was covered in the previous *Handbook*. It's an experiment demonstrating that out-of-equilibrium play drives behavior in the ultimatum game by highlighting the differences in out-of-equilibrium

play with the best shot game which *does* converge to the subgame perfect equilibrium. It is the forerunner to the more general learning models covered in Section III.H.

6.) Fehr, Kirchsteiger, and Riedl (1998): This is a nice summary of the gift exchange literature up to this point in time.

7.) Cox (2004): You need to read at least one paper on the trust game. This is a good one that uses a clever experimental design to start to tease apart the motivations in the game.

8.) Readers choice: There are just too many relevant papers written and yet to be written on regarding preferences. I suggest the reader pick a paper whose results do not square with their intuition and read it carefully. It may just suggest an experiment of your own!

References

Abbink, Klaus, G. Bolton, Karim Sadrieh, and Fang-Fang Tang. 2001. Learning versus punishment in ultimatum bargaining. *Games and Economic Behavior*, 37, 1-25.

Afriat, Sidney N. 1972. Efficiency estimation of production functions. *International Economic Review*, 13, 568-598.

Al-Ubaydli, O., S. Andersen, U. Gneezy, and John A. List. 2006. For love or money? Testing non-pecuniary and pecuniary incentive schemes in a field experiment. Working Paper, University of Chicago.

Akerlof, George. 1982. Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, 97, 543-69.

Andreoni, James and B. Douglas Bernheim. 2008. Social image and the 50-50 norm. *Econometrica* (in press).

Andreoni, James, Castillo, Marco, and Ragan Petrie. 2007. Revealing preferences for fairness in ultimatum bargaining. Working paper.

Andreoni, James and John H. Miller. 2002. Giving according to GARP: An experimental study of rationality and altruism. *Econometrica*, 70, 737-53.

Andreoni, James and Emily Blanchard. 2006. Testing subgame perfection apart from fairness in ultimatum games. *Experimental Economics*, 9, 307-321.

Anderson, L.R., Y.V. Rodgers, and R. R. Rodriguez. 2000. Cultural differences in attitudes toward bargaining. *Economics Letters*, 69, 45-54.

Armantier, Olivier. 2006. Do wealth differences affect fairness considerations? *International Economic Review*, 47, 391-429.

Bardsley, Nicholas. 2008. Dictator game giving: Altruism or artifact? *Experimental Economics*, 11, 122-133.

Baron, D. P. and J. A. Ferejohn. 1989. Bargaining in legislatures. *American Political Science Review*, 83, 1181-1206.

Bellemare, Charles and Bruce Shearer. 2007. Gift exchange within a firm: Evidence from a field experiment. Working Paper, Université Laval, CIRPEE.

Bereby-Meyer, Yoella and Muriel Niederle. 2005. Fairness in bargaining. *Journal of Economic Behavior and Organization*, 56, 173-186.

Berg, Joyce E., John Dickhaut, and Kevin McCabe. 1995. Trust and reciprocity, and social history. *Games and Economic Behavior*, 10, 122-42.

Bewley, Truman, 1998. Why not cut pay? *European Economic Review*, 42, 459-90.

Blanco, Mariana, Dirk Engelmann, and Hans-Theo Normann. 2006. A Within-subject analysis of other-regarding preferences. Working Paper, University of London.

Blount, Sally. 1995. When social outcomes aren't fair: The effect of causal attributions on preference. *Organizational Behavior and Human Decision Processes*, 63, 131-44.

Bolton, Gary E. and Axel Ockenfels. 2006. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *American Economic Review*, 96, 1906-1911.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels 2005. Fair procedures: Evidence from games involving lotteries. *Economic Journal*, 115, 1054-1076.

Bolton, Gary E. and Axel Ockenfels. 2000. ERC: A theory of equity, reciprocity and competition. *American Economic Review*, 90, 166-193.

Bolton, Gary E. and Axel Ockenfels. 1998. An ERC-analysis of the Güth-van Damme game. *Journal of Mathematical Psychology*, 42, 215-226.

Bolton, Gary E. and Rami Zwick. 1995. Anonymity versus punishment in ultimatum bargaining, *Games and Economic Behavior*, 10, 95-121.

Bolton, Gary E. 1991. A comparative model of bargaining: Theory and evidence. *American Economic Review*, 81, 1096-1136.

Brandts, Jordi and Gary Charness. 2000. Hot vs. cold: Sequential responses and preference stability in experimental games. *Experimental Economics*, 2, 227-238.

Brandts, J and Charness, G. 2009. "Does the Strategy Method Induce Different Behavior? A Survey of Laboratory Results," working paper.

Brandts, Jordi and C. Solà. 2001. Reference points and negative reciprocity in simple sequential games. *Games and Economic Behavior*, 36, 138-157.

Brandts, Jordi and Gary Charness. 2003. Truth or consequences: An experiment. *Management Science*, 49, 116-130.

Brosig, J., Weimann J. and C-L Yang**.** 2003. The hot versus cold effect in a simple bargaining experiment. *Experimental Economics*, 6, 75-90.

Brown, M., A. Falk, and Ernst Fehr. 2004. Relational contracts and the nature of market interactions. *Econometrica*, 72, 747-80.

Camerer, Colin F. 2003. *Behavioral Game Theory*. Princeton University Press.

Cameron, Lisa A. 1999. Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry*, 27, 47-59.

Campbell III, Carl M. and Kunal S. Kamlani. 1997. The reasons for wage rigidity: Evidence from a survey of firms. *Quarterly Journal of Economics*, 112, 759 –790.

Casari, Marco and Timothy Cason. 2009. The strategy method lowers measured trustworthy behavior. *Economic Letters* (in press).

Charness, Gary and Mathew Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117, 817-69.

Cherry, Todd L., Peter Frykblom, and Jason F. Shogren.2002. Hardnose the dictator. *American Economic Review*, 92, 1218-1221.

Cohn, A., Ernst Fehr, and L. Goette. 2007. Gift exchange and effort: Evidence from a field experiment. Working Paper. Institute for Empirical Research in Economics, University of Zurich.

Cooper, David J. and Dutcher, E. Glenn. 2009 The Dynamics of Responder Behavior in Ultimatum Games: A Meta-study, Florida State University, working paper.

Cooper, David J. and Carol Stockman. 2002. Fairness and learning in a step-level public goods game. *Games and Economic Behavior*, 41, 26-45.

Cooper, David J. and John Van Huyck. 2003. Evidence on the equivalence of the strategic and extensive form representation of games. *Journal of Economic Theory*, 110, 290-308.

Cooper, David J., Nick Feltovich, Alvin E. Roth, and Rami Zwick. 2003. Relative versus absolute speed of adjustment in strategic environments: Responder behavior in ultimatum games. *Experimental Economics*, 6, 181-207.

Cox, James C. 2004. How to identify trust and reciprocity. *Games and Economic Behavior*, 46, 260-281.

Cox, James C., Daniel Friedman, and Steve Gjerstad. 2007. A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59, 17-45.

Cox, James C., Daniel Friedman, and Vjollca Sadiraj. 2008. Revealed altruism. *Econometrica,* 76, 31-70.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. 2006. Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. Working Paper, University of Illinois.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes. 2006. What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes,* 100, 193-201.

Deci, E.L., R. Koestner, and R.M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627-68.

Duffy, John and Nick Feltovich. 1999. Does observation of others affect learning in strategic environments? An experimental study. *International Journal of Game Theory*, 28, 131-152.

Dufwenberg, M., G. Kirchsteiger. 2004. A theory of sequential reciprocity. *Games and Economic Behavior*, 47, 268-98.

Engelmann, D. and Martin Strobel. 2006. Inequality aversion, efficiency, and maximum preferences in simple distribution experiments: Reply *American Economic Review*, 96, 1918-1923.

Engelmann, D. and Martin Strobel. 2004. Inequality aversion, efficiency, and maximum preferences in simple distribution experiments. *American Economic Review*, 94, 857-869.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. In press. On the nature of fair behavior. *Economic Inquiry*.

Falk, Armin and Urs Fischbacher. 2006. A theory of reciprocity. *Games and Economic Behavior*, 54, 293-315.

Falk, Armin, Ernst Fehr, Urs Fischbacher. 2003. On the nature of fair behavior. *Economic Inquiry*, 41, 20-6.

Falk, Armin and M. Kosfeld. 2006. The hidden costs of control. *American Economic Review*, 96(5), 1611-30.

Fehr, Ernst and Klaus M. Schmidt. 2006. The economics of fairness, reciprocity and altruism – experimental evidence and new theories. *Handbook on the Economics of Giving, Reciprocity and Altruisn*, ed. by. Serge-Christophe Kolm and Jean Mercier Ythier.

Fehr, Ernst, Oliver Hart, and Christian Zehnder. 2008. Contracts as reference points: Experimental evidence. Forschungsinstitut zur Zukunft der Arbeit/ Institute for the Study of Labor, Working Paper # 3889.

Fehr, Ernst and Simon Gächter. 2002. Do incentive contracts undermine voluntary cooperation? University of Zurich, Institute for Empirical Research in Economics, Working Paper #34.

Fehr, Ernst and Armin Falk. 1999. Wage rigidity in a competitive incomplete contract market. *Journal of Political Economy*, 107, 106-34.

Fehr, Ernst and Klaus M. Schmidt. 1999. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114, 817-68.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. Reciprocity as a contract enforcement device. *Econometrica*, 65, 833-60.

Fehr, Ernst, Georg Kirchsteiger, and Arno Reidl. 1993. Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*, 108, 437-60.

Fershtmann, Chaim and Uri Gneezy. 2001. Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, 116, 351-377.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits. 2007. Individual preferences for giving. *American Economic Review*, 97, 1858-1876.

Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6, 347-369.

Fréchette, Guillaume R. 2006. "Session Effects in the Laboratory" CIRANO Working Papers, http://econpapers.repec.org/paper/circirwor/2006s-21.htm.

Fréchette, Guillaume R., John H. Kagel, and Massimo Morelli. 2005. Behavioral identification in coalitional bargaining: An experimental analysis of demand bargaining and alternating offers. *Econometrica*, 73, 1893-1938.

Fréchette, Guillaume R., John H. Kagel, and Massimo Morelli. 2005. Nominal bargaining power, selection protocol and discounting in legislative bargaining. *Journal of Public Economics*, 89, 1497-1517.

Fréchette, Guillaume R., John H. Kagel, and Massimo Morelli. 2008. Pork versus public goods: An experimental study of public good provision within a legislative bargaining framework. Working Paper, New York University.

Fershtmann, Chaim and Uri Gneezy. 2001. Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, 116, 351-77.

Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6, 347-69.

Gale, John, Kenneth Binmore, and Larry Samuelson. 1995. Learning to be imperfect: The ultimatum game. *Games and Economic Behavior*, 8, 56-90.

Gneezy, Uri. 2004. Do high wages lead to high profits? An experimental study of reciprocity using real effort. University of Chicago, Graduate School of Business, Working Paper.

Gneezy, Uri and J. List. 2006. Putting behavioral economics to work: Field evidence of gift exchange. *Econometrica*, 74, 1365-84.

Glaeser, Edward, David Laibson, Jose Scheinkman, and Christine Soutter. 2000. What is social capital? The determinants of trust and trustworthiness. *Quarterly Journal of Economics*,

115, 811-46.

Güth, Werner and Eric Van Damme. 1998. Information, strategic behavior and fairness in ultimatum bargaining: An experimental study. *Journal of Mathematical Psychology*, 42, 227-47.

Hannan, R. Lynn. 2005. The combined effect of wages and firm profit on employee effort. *The Accounting Review*, 80, 167-188.

Hannan, R. Lynn, John Kagel, and Donald Moser. 2002. Partial gift exchange in experimental labor markets: Impact of subject population differences, productivity differences, and effort requests on behavior. *Journal of Labor Economics*, 20, 923-51.

Harrison, Glenn W. and J. Hirshleifer. 1989. An experimental evaluation of weakest link/ best shot models of public goods. *Journal of Public Economy*, 97, 201-225.

Healy, Paul J. 2007. Group reputations, stereotypes, and cooperation in a repeated labor market. *American Economic Review*, 97, 1751-1773.

Henrich, Joseph, et. al., 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28, 795-855.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, 91, 73-78.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 1996. On expectations and monetary stakes in ultimatum games. *International Journal of Game Theory*, 25, 289-301.

Kagel, John H. and Katherine Wolfe. 2001. Tests of fairness models based on equity considerations in a three-person ultimatum game. *Experimental Economics*, 4, 203-19.

Kube, S., M.A. Maréchal, C. Puppe. 2006. Putting reciprocity to work – positive versus negative responses in the field. Working paper.

List, John A. 2007. On the interpretation of giving in dictator games. *Journal of Political Economy*, 115, 482-493.

List, John A. 2006. The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy*, 114, 1-37.

List, John A. and Todd L. Cherry. 2000. Learning to accept in ultimatum games: Evidence from an experimental design that generates low offers. *Experimental Economics*, 3, 11-31.

Montero, Maria. 2007. Inequality aversion may increase inequity. *The Economic Journal*, 117, 192-204.

Oberholzer-Gee, Felix and Reiner Eichenberger. 2008. Fairness in extended dictator game experiments. *B.E. Journal of Economic Analysis & Policy* 8 (1) Art. 16.

Ploner, Matteo and Anthony Ziegelmeyer. 2008. The hidden costs of control: An unsuccessful repetition study. Working Paper, Max Planck Institute of Economics and University of Trento.

Prasnikar, Vesna and Alvin E. Roth. 1992. Considerations of fairness and strategy: Experimental data from sequential games. *Quarterly Journal of Economics*, 107, 865-888.

Rabin, M. 1993. Incorporating fairness into game theory and economics. *American Economic Review*, 83, 1281-302.

Rosenthal, R. and R.L. Rosnow. 1969. Artifact in behavioral research. New York: *Academic Press.*

Rotemberg, Julio J. 2008. Minimally acceptable altruism and the ultimatum game. *Journal of Economics and Organizational Behavior*, 66, 457-76.

Rotemberg, Julio J. 2006. Altruism, reciprocity and cooperation in the workplace. *Handbook on the Economics of Giving, Reciprocity and Altruism*, vol 2, ed. by. Serge-Christophe Kolm and Jean Mercier Ythier, North Holland, 1371-1407.

Roth, Alvin E. and Ido Erev. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8, 164-212.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study. *American Economic Review*, 81, 1068-95.

Shaked, Avner. 2006. On the explanatory value of inequity aversion theory. Working Paper, Bonn University.

Slonim, Robert L. and Alvin E. Roth. 1998. Learning in high stakes ultimatum games: An experiment in the Slovak Republic. *Econometrica*, 66, 569-96.

Tadelis, Steve. 2007. The power of shame and the rationality of trust. Working Paper, University of California, Berkeley, Haas School of Business.

Tversky, A., S. Sattah, and P. Slovic. 1988. Contingent weighting in judgment and choice. *Psychological Review*, 93, 371-384.

Wallis, W. Allen and Milton Friedman. 1942. The empirical derivation of indifference functions. In *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz*, O. Lange, F. McIntyre, and T. O. Yntema, editors, Chicago: University of Chicago Press. 175-189.

Winter, Eyal and Shmuel Zamir. 2005. An experiment with ultimatum bargaining in a changing environment. *The Japanese Economic Review*, 56, 363-385.

Xiao, Erte and Daniel Houser. 2005. Emotion expression in human punishment behavior. Proceedings of the National Academy of Science, 102, 7398-7401.

Xiao, Erte and Daniel Houser. 2008. Avoiding the sharp tongue: Anticipated written messages promoter fair economic exchange. *Journal of Experimental Psychology* (in press).

Table 1
Comparison between transfers in dictator games with and without a lottery present: Zurich results

	Mean transfer (median)	% who keep entire cash endowment	% who play the lottery (dictator game present)	% who play the lottery (dictator game absent)
Standard Dictator Game	2.27 (2.90)	15.4%	--	--
Lottery Treatment	0.38 (0)	39.1%	50.0%	25.8%

Source: Oberholzer-Gee and Eichenberger (2007)

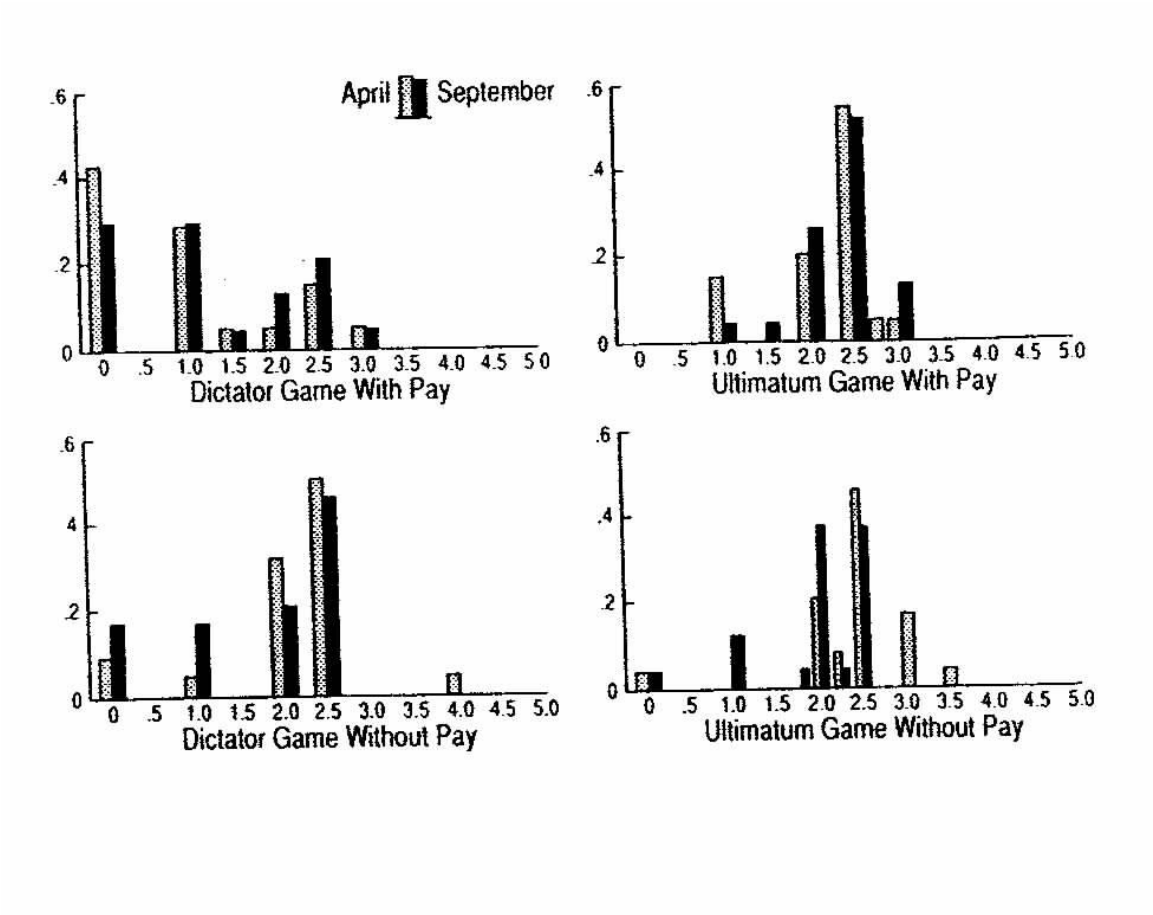
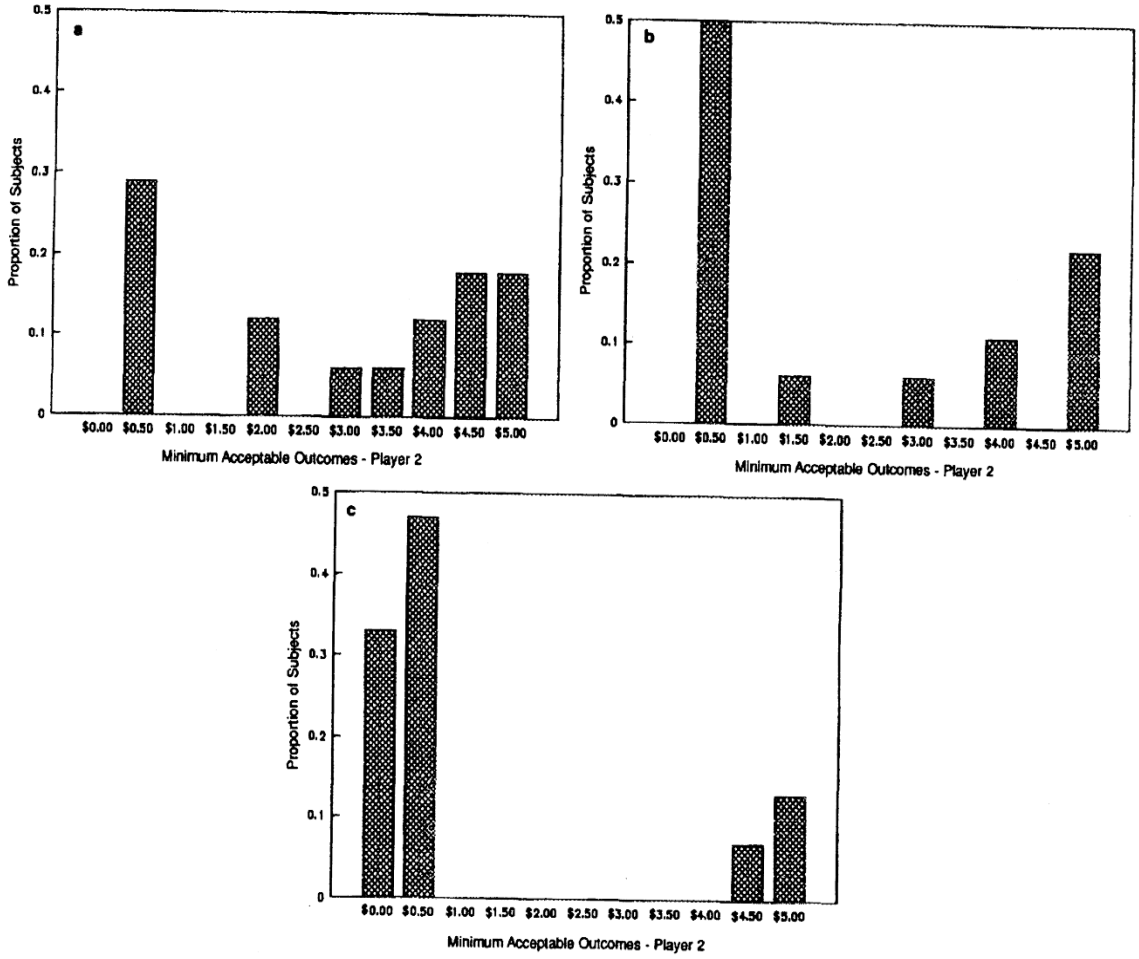


Figure 1: Offers in dictator and ultimatum games, with and without pay. From Forsythe, Horowitz, Savin and Sefton (1994).



Study 1 results. (a) interested party condition, (b) third party condition, (c) random condition.

Figure 2: From Blount (1995).

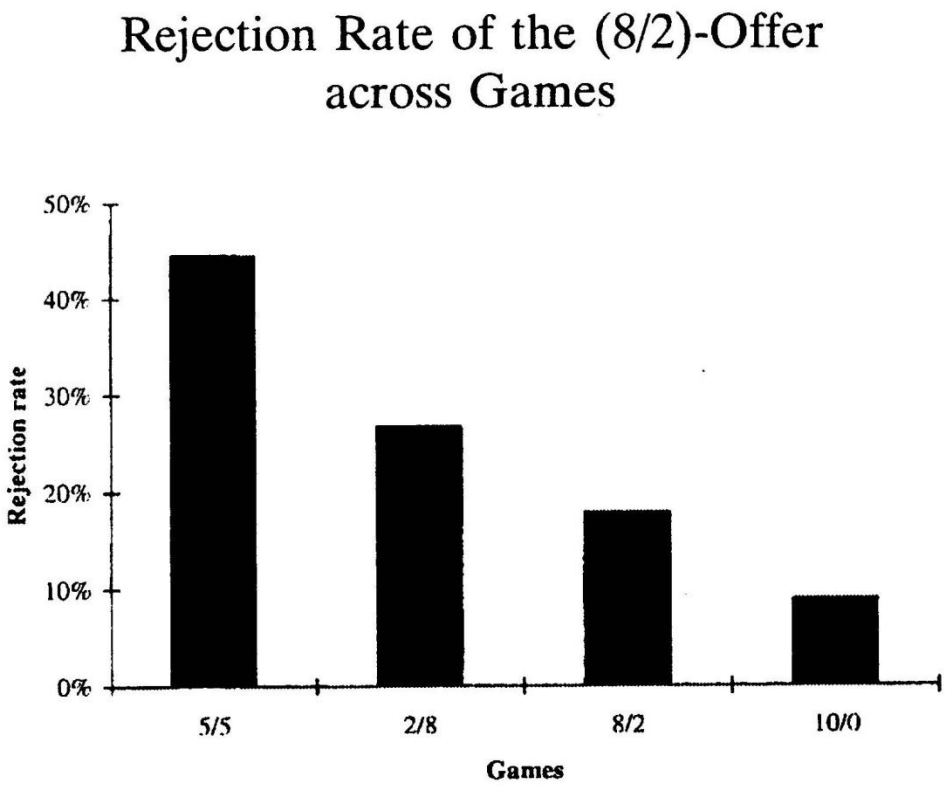
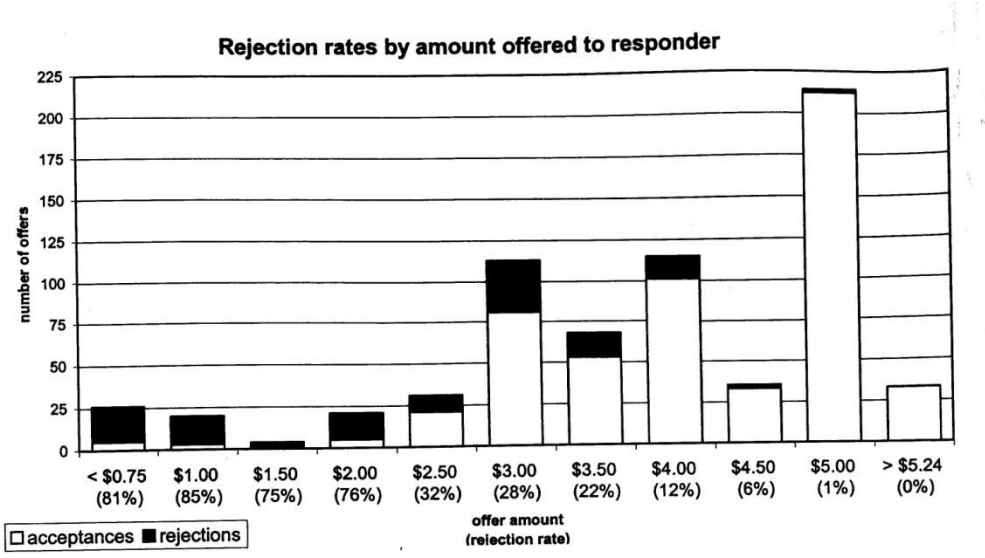
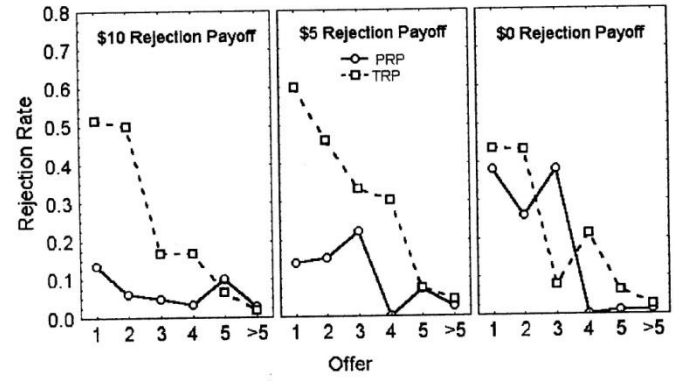


Figure 3: From Falk, Fehr, and Fishbacker (2003).



Non-negative consolation prize sessions.

Figure 4: From Kagel and Wolfe (2001).



PRP:	84	17	7	10	20	12	96	11	6	4	19	14	29	9	7	14	63	28
TRP:	33	16	16	24	42	19	34	29	16	22	41	8	26	27	13	22	48	14

Figure 5: From Bereby-Meyer and Niederle (2005).

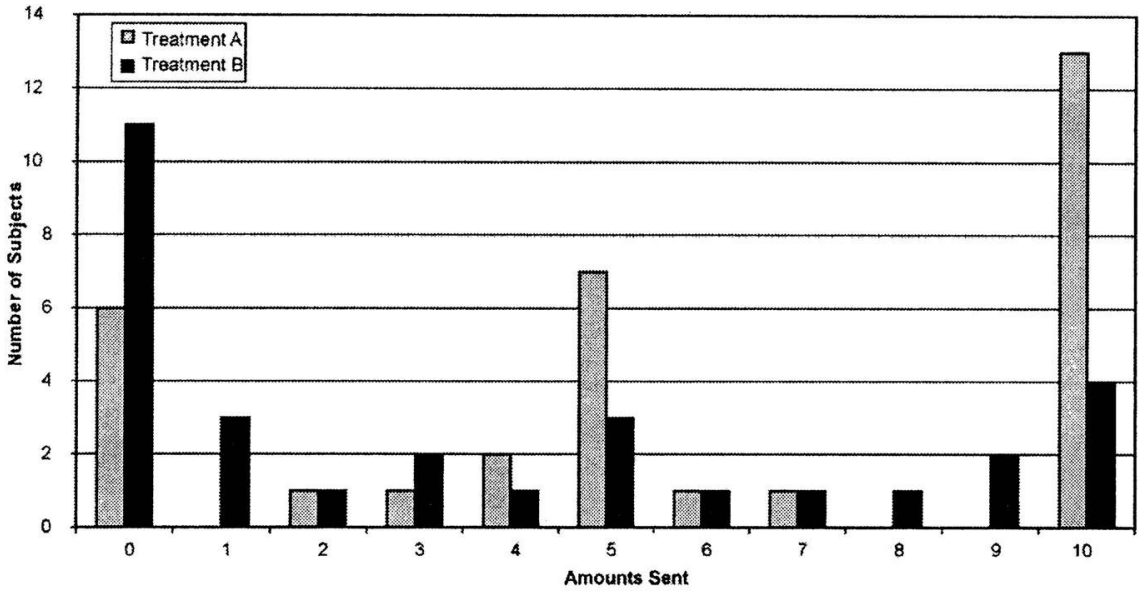
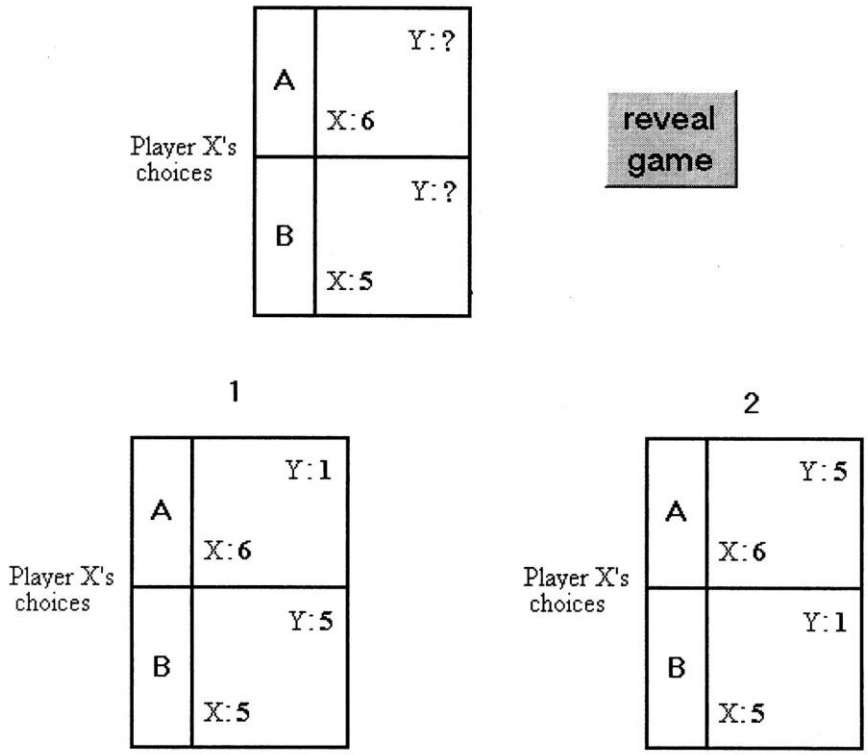


Figure 6: From Cox (2004)



Interface for hidden information treatment

Figure 7: From Dana, Weber and Kuang (2006)

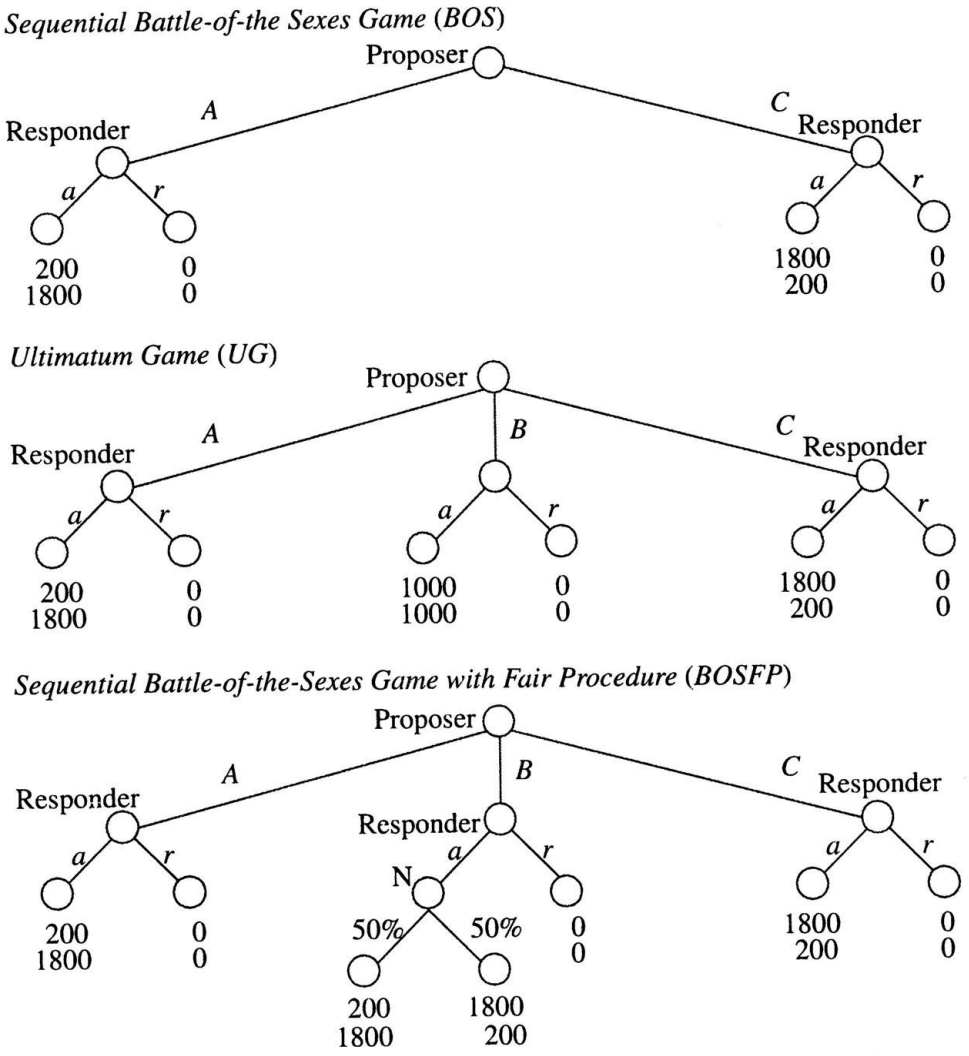
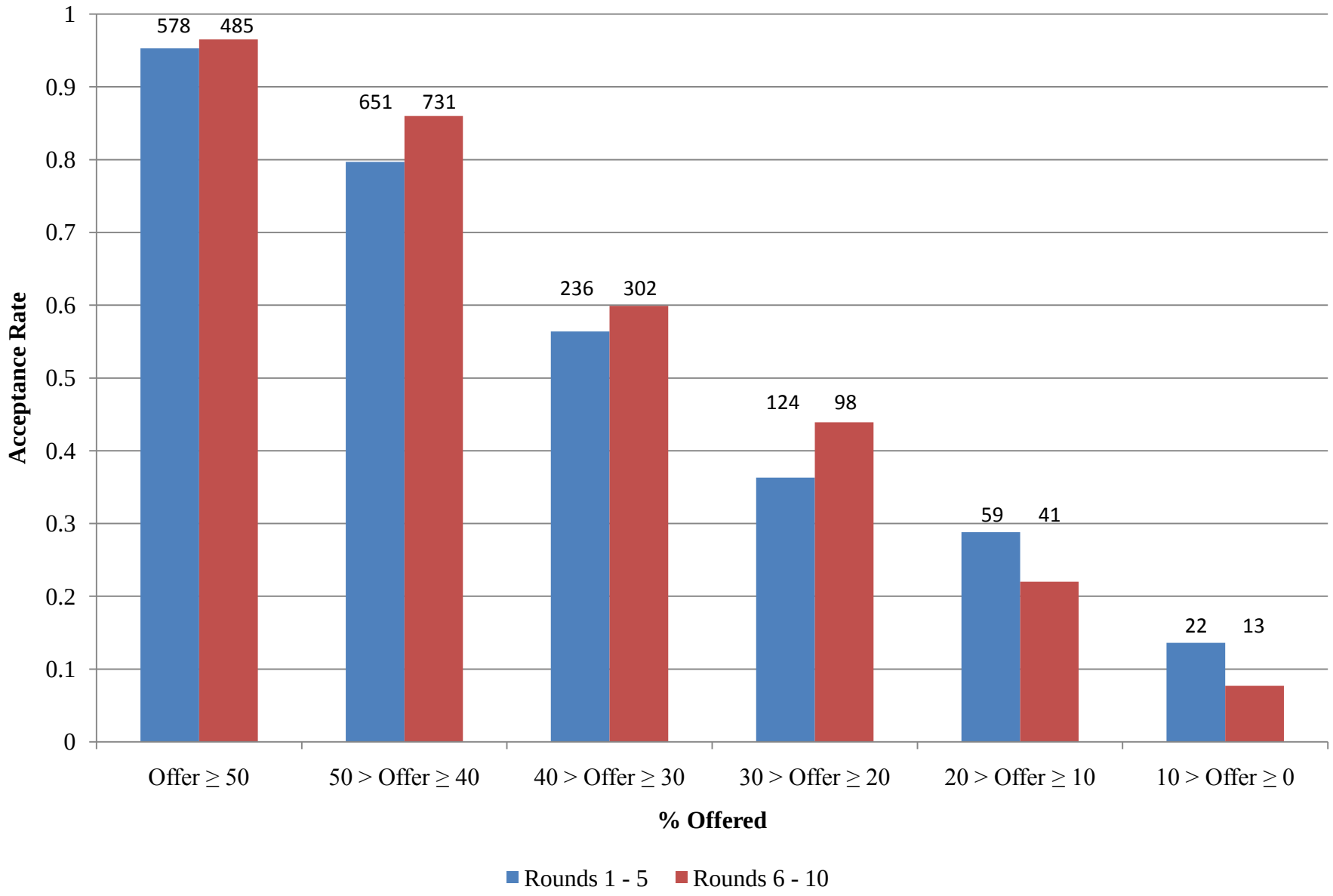


Figure 8: From Bolton, Brandts, and Ockenfels (2005)

Figure 9: Changing Acceptance Rates in Ultimatum Games
(Numbers above the bars give the number of observations for that bar.)



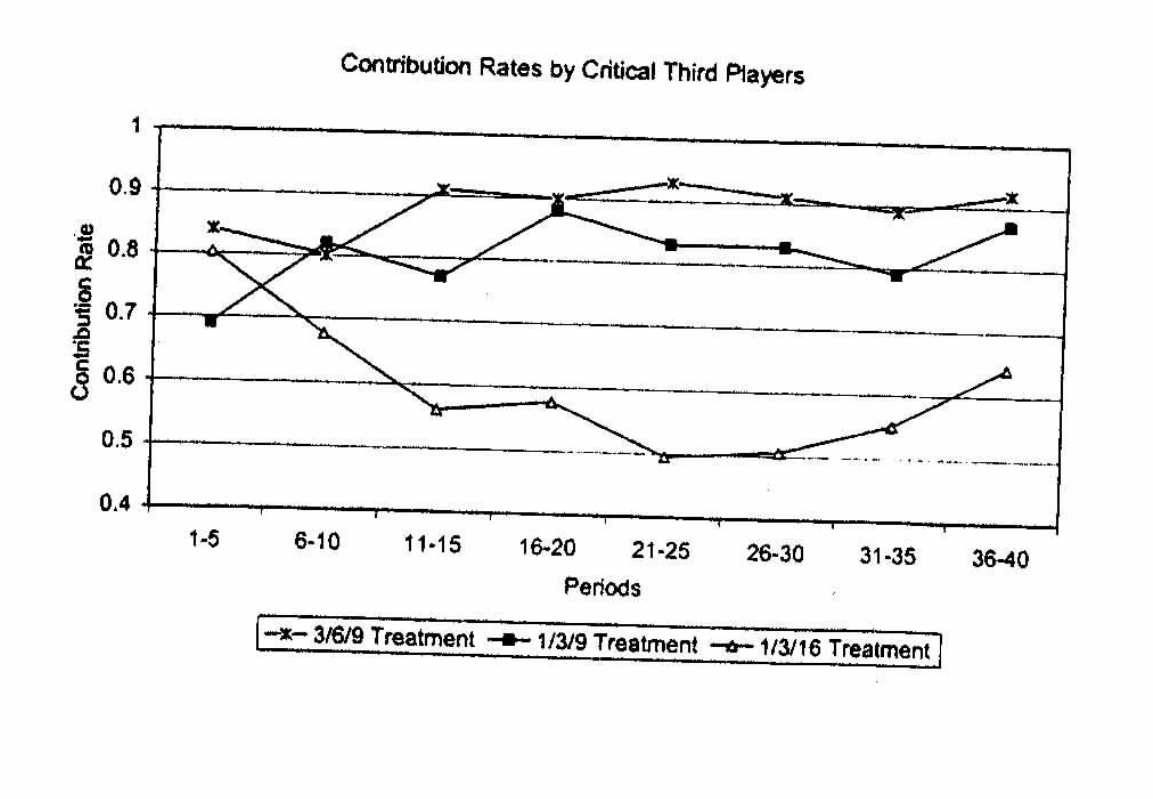


Figure 10: Changes in contribution rates of critical third players in a step-level public goods game. It is always in the self interest of critical third players to contribute to the public good. From Copper and Stockman (2002).

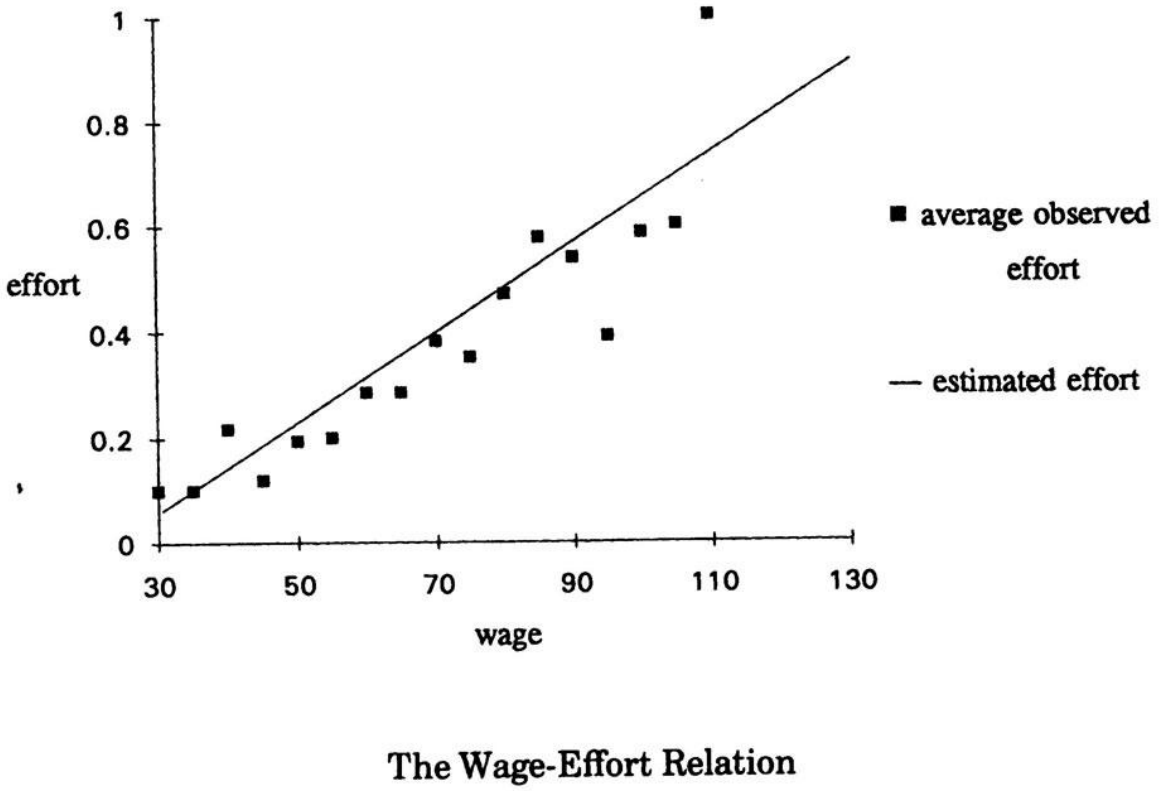
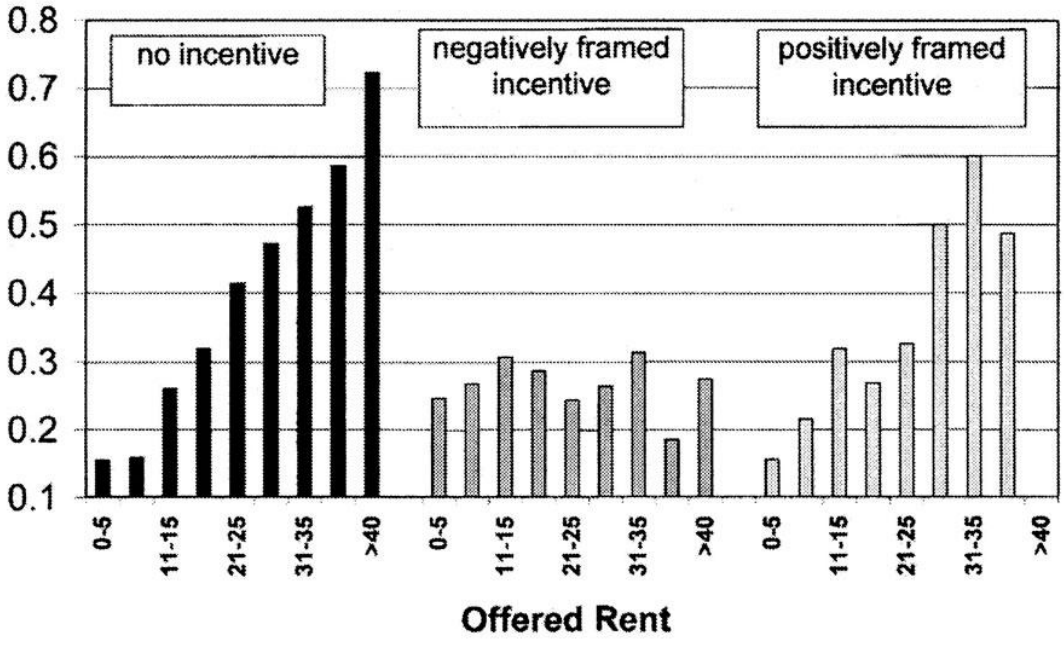


Figure 11: From Fehr, Kirchsteiger, and Riedl (1993)



The impact of explicit incentives on actual average effort (*source*: Fehr and Gächter, 2000b).

Figure 12: From Fehr and Gachter (2002).

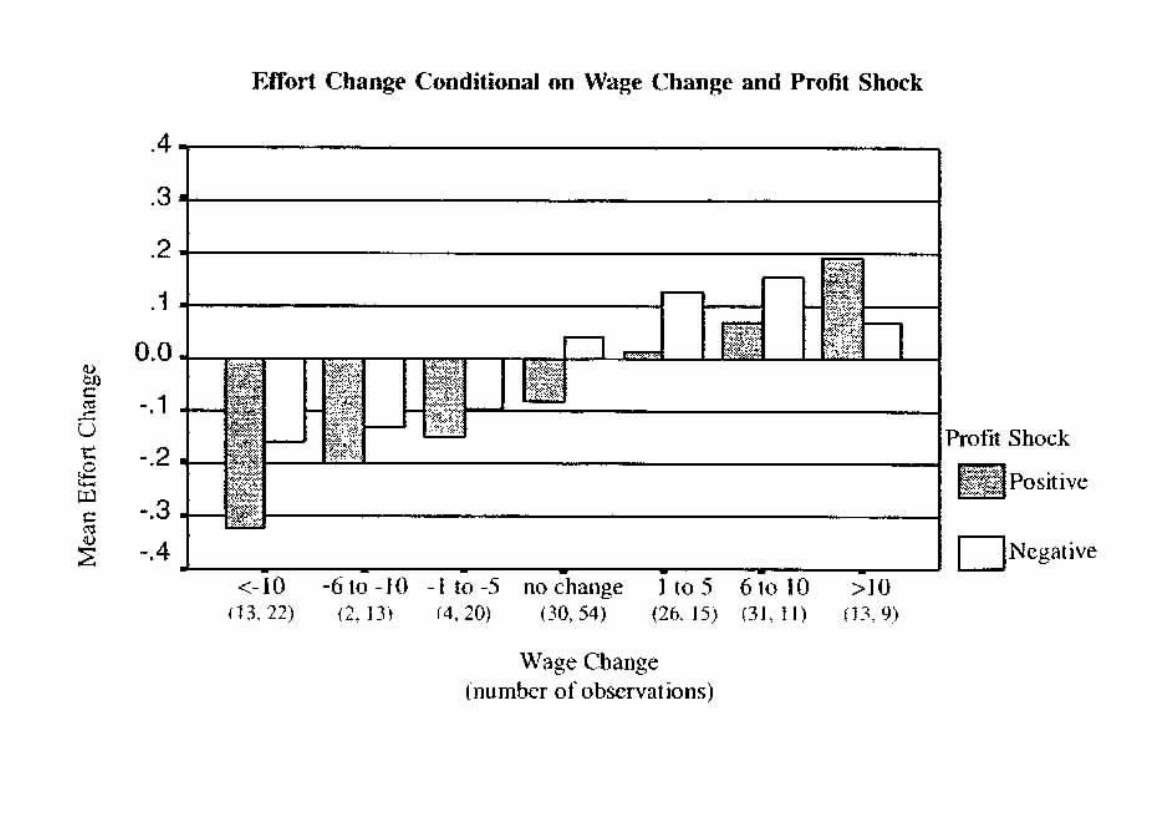
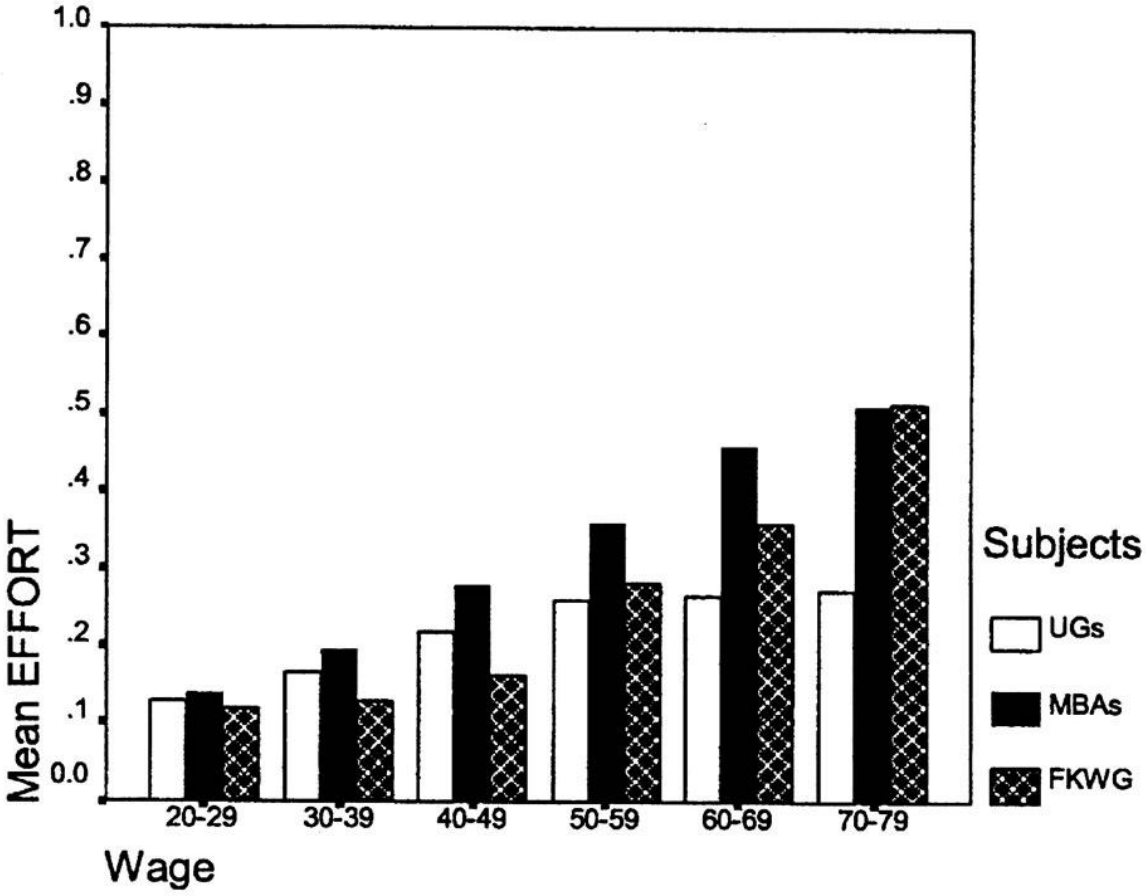


Figure 13: Effort change as a function of wage changes in response to positive and negative profit shocks. From Hannan (2005).



—Mean effort by wage. Wages are truncated at 79 because the data are thin above that level.

Figure 14: From Hannan, Kagel and Moser (2002).

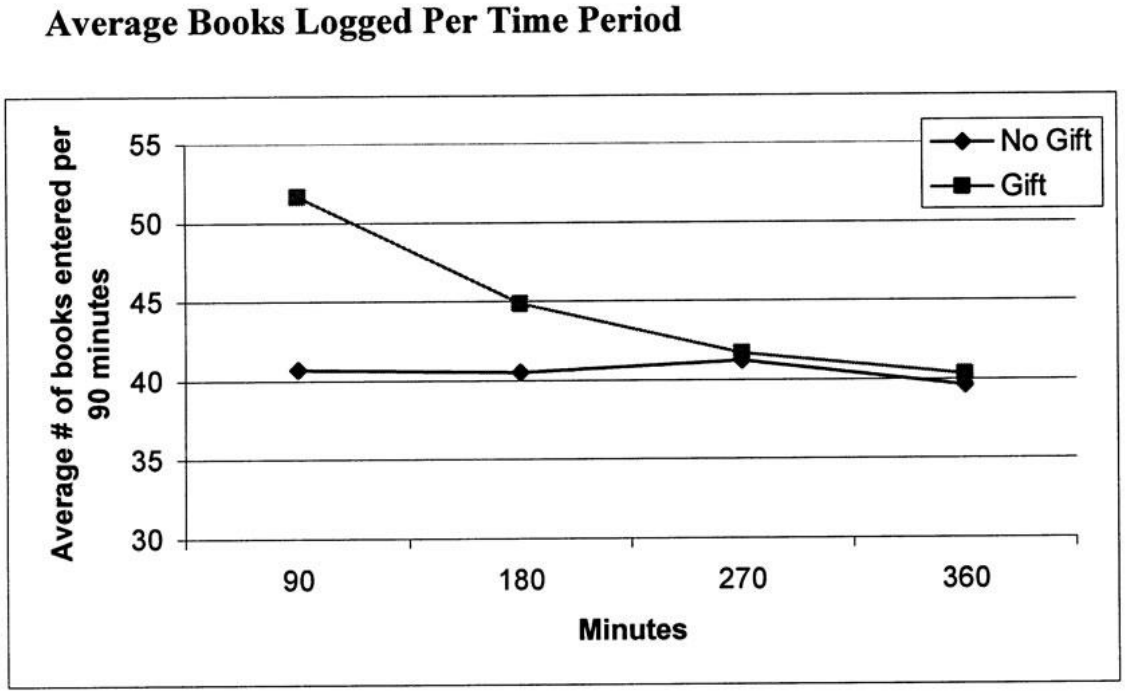


Figure 15: From Gneezy and List (2006)

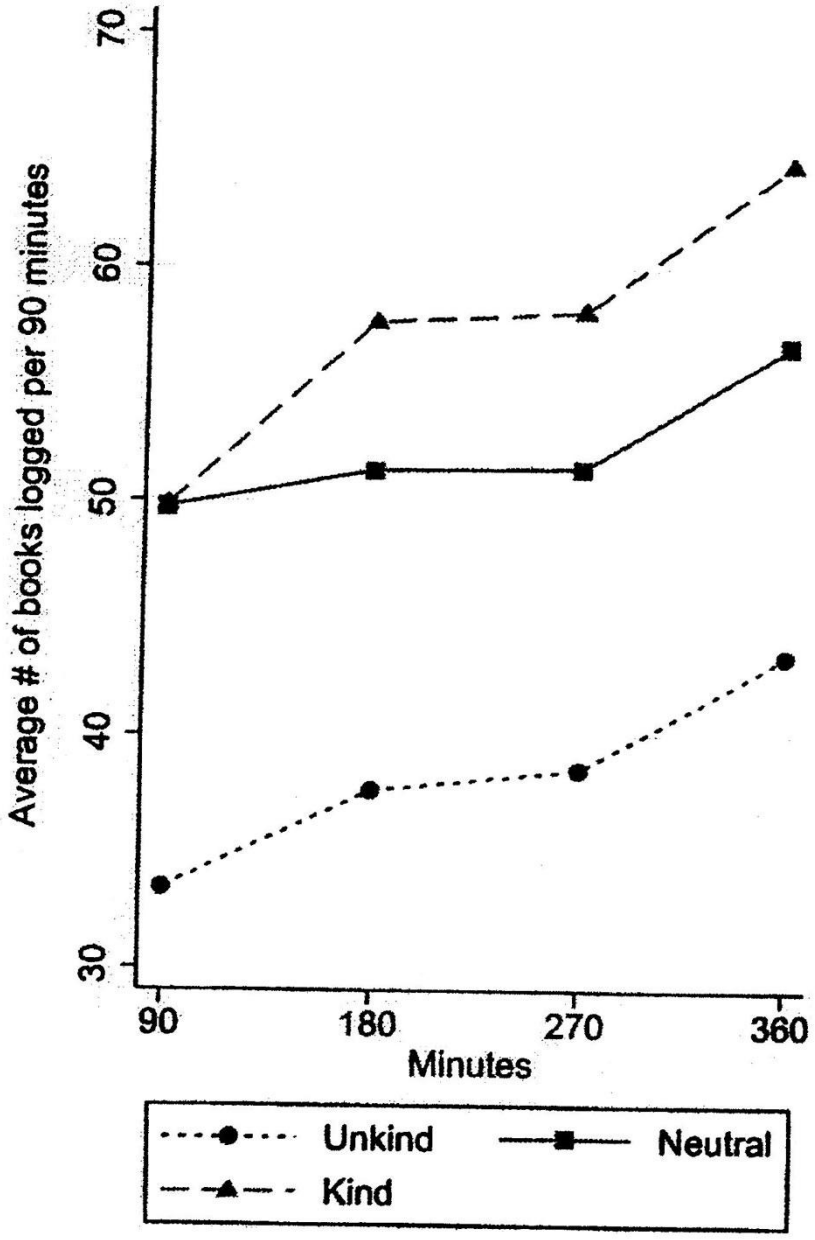


Figure 16: From Kube, Marechal, and Puppe (2006)