Rethinking Reciprocity

Ulrike Malmendier
University of California, Berkeley
malmendier@econ.berkeley.edu,

Vera L. te Velde
University of California, Berkeley
vtevelde@econ.berkeley.edu,

Roberto Weber
University of Zurich
roberto.weber@econ.uzh.ch

October 17, 2013
Contents

1 Introduction ......................................................... 1

2 Understanding Social Preferences ............................... 3
   2.1 One-Sided Giving ........................................... 4
   2.2 Two-Sided Giving .......................................... 5
   2.3 The Missing Piece .......................................... 9

3 Experiment ......................................................... 9
   3.1 Basic Idea .................................................. 10
   3.2 Experimental Design ....................................... 11
   3.3 Experimental Results ...................................... 13

4 Discussion ......................................................... 17

5 Model and Estimation of Extrinsic Motives ................... 20
   5.1 Model ....................................................... 21
   5.2 Estimation .................................................. 23

6 Conclusion ......................................................... 28

7 Tables ............................................................. 37

8 Figures ............................................................ 45
Abstract

Reciprocal behavioral has been found to play a significant role in explaining market outcomes in labor, health, and public economics as well as political economy. Despite mounting empirical evidence, economists are still struggling to converge on the correct model of the underlying motives. Existing theories posit intrinsic preferences for the welfare of others, equality aversion, or utility from repaying the kindness of others. More recent evidence reveals that seemingly altruistic people are often reluctant to share and try to avoid giving opportunities. Extrinsic factors such as social image, self image, or social pressure appear to be important determinants of sharing in ‘one-sided’ (no-reciprocity) environments. However, this revision of conventional theoretical motives for sharing has had little spillover to ‘two-sided’ reciprocity environments, where one individual responds to the actions of another.

We review the literature on reciprocity and point to the relative lack of attention paid to extrinsic factors. We then present a novel experiment that explores the importance of intrinsic versus extrinsic factors in driving reciprocal behavior. We find that, in a laboratory reciprocity setting (the double-dictator game), failure to account for extrinsic motives leads to a significant overestimation of intrinsic motives such as fairness and altruism. We use the experimental data to illustrate the importance of combining reduced-form and structural analyses in disentangling intrinsic and extrinsic determinants of pro-social behavior.

Keywords: Social preferences, social image, social pressure, extrinsic and intrinsic motivations, structural estimation, double-dictator game.
1 Introduction

Reciprocal behavior is of increasing interest in many areas of economics. Labor economists argue that firms pay above-market wages in order to induce reciprocal behavior among their workers (Akerlof 1982, Bewley 2009, Fehr et al. 2009), which in turn might explain involuntary unemployment (Akerlof and Yellen 1988, Akerlof and Yellen 1990) and inflation (Okun 1981). Conversely, workers who feel mistreated by their employer exert negative reciprocity and produce faulty products (Krueger and Mas 2004). Public economists show that inducing reciprocity with small gifts helps to raise funds for public projects, such as national parks (Alpizar et al. 2008), or for charities (Falk 2007), and argue that reciprocity is important for tax compliance (Feld and Frey 2007). Political economists provide evidence that politicians target pre-election transfers towards reciprocal individuals to increase their vote share (Finan and Schechter 2012). And health economists show that pharmaceutical representatives exploit the reciprocity of doctors with small gifts (Brennan et al. 2006). This literature in economics builds on earlier work in psychology, including the seminal work by Cialdini on the use of small gifts in marketing and politics with the aim to trigger reciprocity (see, e.g., Cialdini (1993)).

The recognition of reciprocity as a determinant of economic behavior extends the relevance of social preferences beyond their explanatory power for charitable giving and other ‘one-sided’ acts of sharing and giving to a broader class of market interactions. It suggests that market features such as dynamic pricing strategies (e.g., low introductory prices), consumer loyalty, or employee turnover are affected by social preferences. Profit-maximizing firms need to anticipate and respond to the non-standard behavior of their consumers.

Despite the empirical importance of reciprocal behavior, economists still struggle to converge on the best model of reciprocity. Existing theories rely on the response to good intentions or the kindness of others as the trigger for reciprocal behavior, e.g. Rabin (1993) or Levine (1998). Under these theories, people reciprocate because another person’s kind act or benevolent nature increase the intrinsic utility of acting kindly toward this person.

In this article, we propose that the analysis of reciprocity needs to be re-thought. Our argument is motivated by recent evidence on the motives underlying a simpler, ‘one-sided’ type of pro-social behavior. Recent literature shows that many seemingly altruistic people are reluctant to share and avoid the opportunity to share if they can (Dana et al. 2006, Broberg et al. 2007, Lazear et al. 2012, DellaVigna et al. 2012). Such evidence has led to a revision of the conventional theoretical motives thought to underly generous or charitable behavior. In addition to an intrinsic preference for equality or social welfare, extrinsic factors such social image, self image, or the response to social pressure or norms appear to be important determinants of seemingly altruistic behavior (Akerlof and Kranton 2000, Bénabou and Tirole 2006, Battigalli and Dufwenberg 2007, Harbaugh 1998). These advances in our understanding of ‘one-sided’ giving decisions have thus far had little spillover to our understanding of ‘two-sided’ pro-social behavior, where the decision to give is a function of the earlier generous behavior of the other person. Our review discusses the existing literature on reciprocity, including the leading theoretical models, and points to the common failure to account for extrinsic motives.
We then present a novel experiment that explores the importance of intrinsic versus extrinsic determinants of reciprocal behavior. We employ a double-dictator game, which allows us to compare giving in a positive- and a negative-reciprocity setting with giving in a neutral setting (the standard dictator game). We modify those games by giving dictators the option to avoid the sharing decision. The 3x2 design (neutral/positive/negative x no-sorting/sorting) allows us to disentangle intrinsic and extrinsic factors determining giving in one-sided versus two-sided (reciprocity) giving environments.

The findings cast a new light on the motives underlying reciprocity. First, when the other party induces positive reciprocity (with prior sharing), giving increases relative to the baseline (standard) dictator game, consistent with prior findings. The increase persists after we introduce sorting, relative to a baseline dictator game with sorting. However, comparing double-dictator games with and without sorting, we also continue to find a significant decrease in giving under sorting, very similar in size to the sorting effect in the single-dictator game. Thus, positive reciprocity does not eliminate reluctance to share and we can conclude that extrinsic factors are important determinants of sharing in reciprocity environments as well. We also find that negative reciprocity, induced by receiving $0 out of $2 in the first stage of the double-dictator game, virtually eliminates giving in the setting with sorting. It even induces some people to sort in and then share zero. The reduced-form results suggest that failure to account for extrinsic motives leads to a significant overestimation of intrinsic motives such as fairness and altruism in positive-reciprocity settings, and to a significant underestimation of spite (negative altruism) in negative reciprocity settings.

We confirm these conclusions employing a simple structural model which distinguishes between intrinsic and extrinsic motives. Even after the inducement of positive reciprocity, we estimate a significant influence of extrinsic factors, similar in size to the strength of external factors in one-sided (standard) dictator games.

We also compare our estimates to those based on a naive structural model that does not account for extrinsic motives. Consistent with the reduced-form results, we find that failure to account for extrinsic motives leads to a sizable overestimation of intrinsic motives, both under positive and under negative reciprocity. At the same time, however, the estimation results also show that the amount of “additional” intrinsic motivation, in the sense of “reciprocity-induced” altruism or similar intrinsic motives under positive reciprocity is estimated approximately correctly in the naive model. In other words, it is correct to attribute the additional giving in a positive-reciprocity environment, compared to a neutral setting, to intrinsic preferences, and even a naive model delivers the correct estimate. This robustness reflects the fact that extrinsic motives are remarkably stable across both environments. This conclusion is somewhat less true for the negative-reciprocity environment: The decrease in intrinsic willingness to share after an unkind treatment, relative to a neutral treatment, is slightly underestimated when neglecting extrinsic factors and is strongly overestimated when allowing only for a reciprocity-insensitive extrinsic motive. In other words, failure to account for extrinsic motives and their context-dependence is particularly detrimental when estimating the motives for and extent of negative-reciprocal behavior.

In summary, the experimental data confirms that extrinsic motives are a significant de-
terminant of pro-social behavior in reciprocity environments, similar in strength to their role in ‘one-sided’ environments without reciprocity. This insight is relevant to the economic analysis of reciprocal behavior in markets as it suggests that market participants aiming to ‘trigger’ positive reciprocity need to account for observability and avoidance options. For example, a charity might benefit from giving a small gift to the potential donor, as shown by Falk (2007); but the gift will be most effective if the recipient cannot avoid the subsequent request for a donation.

A second main insight from the experimental data relates to the value of combining reduced-form and structural analyses in disentangling determinants of pro-social behavior when the outcome (sharing) could be explained by rather different underlying psychological motives. The structural estimation allows us to decompose the share of giving that is due to intrinsic versus extrinsic determinants, both in one-sided and in two-sided giving contexts.

We would like to emphasize that the intrinsic determinants of reciprocity need re-thinking beyond the confounds with extrinsic factors. Even motives that are cleanly identified as intrinsic, i.e., the giver seeks a giving opportunity rather than avoids it, are not necessarily ‘altruism’ or ‘fairness.’ For example, rather than caring about the intentions or kindness of the other person (Rabin 1993, Levine 1998), a giver might feel an obligation to reciprocate (Cialdini 1993, Sugden 1984). For example, Malmendier and Schmidt (2012) observe that reciprocal behavior is triggered even if the preceding “gift” was solely motivated by the selfish goal of inducing a reciprocal response, which is inconsistent with an intention-based or type-based view of reciprocity, but consistent with an obligation- or norm-based understanding (see also (Postlewaite 2011, Krupka et al. 2012, Bicchieri 2006)). We will not explore this distinction here.

The remainder of the paper is organized as follows. In Section 2, we review the literature on social-preferences, both in ‘one-sided’ and in ‘two-sided’ giving contexts. We suggest that recent insights on extrinsic determinants of giving in one-sided situations need to be applied to the study of reciprocity and other social preferences in two-sided giving contexts. We then present the results of a novel experiment that allows us to compare giving in a neutral versus a positive-reciprocity and a negative-reciprocity environment (Section 3). We then discuss a simple model integrating intrinsic and extrinsic factors in the analysis of reciprocity and structurally estimate their relative importance (Section 4). Section 5 concludes.

2 Understanding Social Preferences

The relevance of social preferences has been a source of debate in the profession for several decades. How do we explain sharing behavior in standard economic settings when it is inconsistent with a classical model of self-interested preferences? Within this broad area of inquiry, social preferences, most attention has been paid to what one might call ‘one-sided’ acts of kindness or generosity, where one party shares with another party without consideration of the other party’s sharing behavior. This article focuses on ‘two-sided’ pro-social behavior, where sharing is a response to the other person’s pro-social behavior. However,
since our motivation for rethinking the reciprocity literature stems from advances in the literature on one-sided sharing decisions, we first review the two waves of literature focusing on one-sided sharing situations.

2.1 One-Sided Giving

A large volume of laboratory and field experiments documents voluntarily sharing behavior, and studies the individual characteristics and contextual factors that influence such sharing. These studies show that generosity varies based on determinants such as the gender of decision makers or other personal characteristics (Andreoni and Vesterlund 2001, Andreoni and Miller 2002, DellaVigna et al. 2013), the framing of the decision as “giving” or “taking,” “sharing” or “dividing,” and the source of the surplus, as well as factors such as risk, community size, culture as well as the social context in which an altruistic choice is embedded (Hoffman et al. 1996, Cherry et al. 2002, Fong and Luttmer 2009, Henrich et al. 2010, Brock et al. 2013). While sharing can vary substantially based on these factors, the existing research provides robust evidence of a significant willingness to share with others that is inconsistent with purely self-interested preferences.

In response to these findings, economists have developed simple models of preferences that can account for such behavior. Leading models include altruism in the form of utilitarianism (Andreoni and Miller 2002) or in the form of maximin preferences (Charness and Rabin 2002), as well as different specifications of inequality aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000). That is, the assumption became that people derive utility from implementing equal or fair outcomes.

More recent research has demonstrated that the motives underlying a sharing decision are more complex. Several studies document a puzzling phenomenon: people share voluntarily when asked to decide between sharing and not sharing; but most of them prefer to avoid the decision altogether and keep their endowment (Dana et al. 2006, Broberg et al. 2007). The novel design feature in these studies is to give the decision maker the option of avoiding the decision to give altogether. For example, in the laboratory study of Lazear et al. (2012), subjects are allowed to “opt out” of playing the dictator game. Or, in the field study of DellaVigna et al. (2012), the exact timing of a door-to-door fundraiser is pre-announced so that people can choose not to open the door of their home. In both cases, people can avoid the socially displeasurable act of having to say “no” when presented with a sharing opportunity.

Along similar lines, contextual features that allow decision makers to obscure the relation between their behavior and unfair outcomes also decrease the willingness to act pro-socially. For example, Dana et al. (2007) show that giving decreases when the recipient does not know for sure how a choice was made, or when the dictator does not know how his choice will affect the recipient.\footnote{A more extreme way to remove transparency is having the recipient unaware of a game entirely. Dana et al. (2006) show that dictators are less generous in this case. Koch and Normann (2008) and Johannesson (2000), instead, find similar generosity in a standard dictator game and a variant with uninformed recipients. It remains to be shown whether differences in their design (double blind setup, recipients randomly chosen} In fact, dictators prefer not to find out about the effect of their
choice. Andreoni and Bernheim (2009) manipulate both anonymity and the probability that a computer secretly overrides the dictator's choice with a known default, showing both that anonymity reduces sharing and that dictators will take the opportunity to hide their selfish actions when possible. Grossman (2012) uses a similar probabilistic dictator game to demonstrate the importance of social signaling relative to self-signaling. Moreover, several of the experiments mentioned above have shown that people prefer to avoid letting the recipient know about the game, even at a cost (Dana et al. 2006, Broberg et al. 2007, Lazear et al. 2012).

The importance of avoidance options can also be seen in earlier experiments that manipulate anonymity and observability. Hoffman et al. (1994) and Hoffman et al. (1996) show that dictator game giving drops when using a double-blind framework. Conversely, Bohnet and Frey (1999b) and Bohnet and Frey (1999a) show that giving increases when dictator and recipient face each other. Relatedly, Franzen and Pointner (2012) show that dictator-game sharing is reduced when choices are concealed from the experimenter using a randomized response technique.

Anonymity and observability appear to play a similar role outside of the lab. For example, Mas and Moretti (2009) show that store workers work harder when observed by other workers who work harder. Gerber et al. (2008) find that revealing voting behavior to neighbors increases voter turnout. And DellaVigna et al. (2012) show that face-to-face interaction with a solicitor creates social pressure to donate to charity. In other words, generosity decreases significantly when individuals can avoid “saying no” directly to someone or being observed by someone.

These studies had a significant impact on the way economists think about social preferences. They revealed that people who share do not necessarily derive utility from sharing in the manner assumed by early social preference models. Instead, the motive driving much voluntary sharing appears to be a concern to not seem selfish or greedy. Rather than valuing the opportunity to act generously or altruistically, people are often motivated by an aversion to the guilt or shame that comes from disappointing the expectations of others, violating norms of sharing, or giving others the impression that one is selfish (Bénabou and Tirole 2006, Battigalli and Dufwenberg 2007, Andreoni and Bernheim 2009).

To sum up, the above stream of social-preferences research can be organized as follows: first, empirical evidence that people act in a manner consistent with a preference for altruism or fairness and theoretical models of such a motivation; second, evidence that such behavior is susceptible to variations in observability and manipulations that allow the maintenance of a positive social image while acting selfishly, along with a new class of social-preference theories that account for such extrinsic concerns.

### 2.2 Two-Sided Giving

So far, the new insights have largely failed to spill over to a broader set of social preferences beyond the ‘one-sided’ individual decision-making. Here, we focus on one widely-discussed
type of social preference, reciprocity. Reciprocity is the tendency to reciprocate kind acts with kindness and unkind acts with spite (Rabin 1993, Charness and Rabin 2002, Falk and Fischbacher 2006). A preference for acting reciprocally has been advanced as underlying many puzzling phenomena, both in the field, as discussed in the introduction, and in the laboratory. In what follows, we provide a brief review of some of the extensive literature on reciprocity, including a description of theoretical accounts for the phenomenon.

Evidence of reciprocity in the laboratory is widespread and robust (see also Fehr and Gächter (2000b) for a review). The pattern of reciprocal behavior survives in one-shot, anonymous scenarios (Hoffman et al. 1998), persists over time and through learning opportunities (Keser and van Winden 2000), and people even reciprocate on behalf of others (Carpenter and Matthews 2004, Carpenter et al. 2004, Fehr and Fischbacher 2004), a phenomenon referred to as “indirect” or “social” reciprocity.

Positive reciprocity is observed as the response to costly investments by others in trust games, often strongly enough to yield a non-negative return to kindness (Berg et al. 1995). Pillutla et al. (2003) show that this reciprocal impulse is strengthened when the trustor takes a very unselfish, risky action to begin with, indicating that outcomes are not the only driving force behind reciprocity.

Negative reciprocity is observed as costly punishment in public-good games (Fehr and Gächter 2000a, Ostrom et al. 1992). Croson (2007) uses variants of the public-good game to show that reciprocity is a stronger motivation than either altruism or the moral commitment to act in the manner one would prefer everyone to act.

Evidence from ultimatum game experiments, using populations from many countries, shows that reciprocity is a cross-cultural phenomenon that persists even when the stakes are as high as several months' wages (Henrich et al. 2001, Roth and Erev 1995, Camerer and Thaler 1995). A critical element driving reciprocity in such games seems to be the intention and agency behind the initial act (Blount 1995, Brandts and Solà 2001). Thus, in the ultimatum game, reciprocity manifests as costly punishment for “bad” behavior, which is evaluated based on more than just the resulting outcomes.

Closely related to trust and ultimatum games are findings on gift-exchange, both in the laboratory and in the field. For example, a laboratory experiment by Fehr et al. (1997) finds that employers who offer high wages are rewarded with higher effort, even when wages are fixed at a flat rate and effort is non-contractible. Using a similar design, Fehr et al. (1993) show that buyers who offer a high price for a good are rewarded with higher quality.

Field experiments often find weaker evidence of gift exchange than the laboratory experiments, but confirm its existence and help to identify the conditions under which gift-exchange reliably occurs. For example, Gneezy and List (2006) find that workers reciprocate surprise gifts with higher effort, though the effect does not persist over time (see also List (2006)). Kube et al. (2012) show that the type of gift is important: small non-monetary gifts and monetary gifts presented thoughtfully are better at inducing positive reciprocity.

The field data also confirms both the positive and the negative side of reciprocal behavior. On the side of positive reciprocity, Falk (2007) shows that donation requests are returned more frequently when they include a token gift (a postcard). On the side of negative reci-
procity. Greenberg (1990) documents a rise in employee theft after wage cuts, and Krueger and Mas (2004) show that employee sabotage during a period of disputes with management was responsible for a dramatic increase in defective tires. Kube et al. (2011) argue that negative reciprocity—induced by a surprise wage cut—is stronger and more robust than positive reciprocity in field settings. (See also Baumeister et al. (2001).)

Models of reciprocity have naturally developed alongside these experimental results. Most fall in three broad classes: outcome-based, type-based (a.k.a. interdependent preferences), and intentions-based models. We discuss these models here with an eye toward our main argument, the relevance of extrinsic determinants of reciprocal behavior, and the particular experimental analysis of a one-shot, non-strategic, anonymous game, to be discussed in the next two sections.

In the category of outcome-based models, one type of model that can account for reciprocity is pure altruism. Altruists incorporate others’ material outcomes into their utility just like any other good to which they wish to allocate wealth. This approach has been employed by many authors since Becker (1961). Bergstrom et al. (1986), for example, use this type of model to explain the private provision of public goods. Andreoni and Miller (2002) formulate a model of utilitarianism which has been widely used in the social-preferences literature. Impure altruism, also known as “warm glow” (Andreoni 1989, Andreoni 1990), works similarly in that a giver obtains direct utility from the another person’s material outcome. In this case the utility stems from the act of giving, even though the giver does not care directly about the outcome of the recipient (e.g., the total donations obtained, including from other sources).

Another class of outcome-based models that can generate reciprocal behavior assume distributional preferences. One example are maximin preferences as in Charness and Rabin (2002), where a giver would like the minimum payoff of any person to be as large as possible. Another example is inequality aversion, as modeled by Fehr and Schmidt (1999) or Bolton and Ockenfels (2000). All of these models assert that we care about others’ outcomes insofar as we care about our own outcomes relative to those of others. Related models incorporate absolute norms of fairness, such as suggested by Ledyard (1997) and used by Cappelen et al. (2007).

Outcome-based models of distributional preferences predict reciprocal behavior as a byproduct of individuals trading off their personal material outcomes and fair outcomes. For example, a kind act by one party (e.g., a firm paying a high fixed wage) may create an inequality that the other party then seeks to mitigate through a reciprocal kind act. That is, reciprocity is implied by agents trying to re-balance their allocations in response to kind or unkind transfer from the others.

In the category of type-based models, or interdependent preferences, the seminal model of Levine (1998) allows the weight someone places on his partner’s material outcome to depend both on a personal altruism parameter and the partner’s altruism parameter. That

---

2Sobel (2005) and Cooper and Kagel (forthcoming) provide excellent additional discussions of reciprocity.

3Charness and Rabin (2002) also allow an agent to place a lower weight on others’ payoffs when they have “misbehaved,” thus incorporating an element of the type-based models discussed next.
is, nice people are nicer, and everyone is nicer to nice people. On the opposite end of the spectrum, the range of altruism parameters also allows for spite. Rotemberg (2008) develops a variant of this model, with the twist that decision makers treat others according to some default altruism parameter as long as their actions meet a “minimally acceptable altruism” threshold, but switch to a more spiteful altruism parameter if the other person’s actions fall below that threshold.

These models treat reciprocity as the product of a signaling game. Actions signal a person’s kindness, and this affects others’ altruistic preferences toward that person. The type-based approach also accounts for indirect reciprocity, i.e., treating a person according to how they have previously treated others.

In the category of intention-based models, the foundational example is Rabin’s (1993) psychological equilibrium model, in which beliefs enter directly into utility. Players maximize a utility function that has both material utility and reciprocal kindness (or spite). The kindness (or spite) of an action is defined not only by what other actions are available to an agent, but also by the agent’s beliefs about what the other player will do, insofar as these beliefs reveal the agent’s intentions. If a player behaves neutrally, kindness and spite are absent, and the second player simply maximizes his material outcome. However, if the first player is kind, by taking an action that, based on what she believes the second player will do, makes the second player better off, then the second player may prefer to reciprocate those intentions. In equilibrium, each player’s beliefs about the others’ actions (and second- and third-order beliefs) must be correct, and players must optimally respond to those actions and to the “kindness” that is implied by those actions.

A key result of this approach is the existence of additional equilibria, relative to a game in which players consider only their own material payoffs. Here, players may act mutually kindly or mutually spitefully toward one another. This model diverges from models such as Fehr and Schmidt (1999) and Levine (1998) in that people care directly about psychological phenomena, rather than their preferences over everyone’s consumption being changed by psychological phenomena. More concretely, people directly care about treating others the way they expect these others to treat them, and not solely about the resulting material outcomes. For example, in contrast with inequality aversion, the second mover is inclined to reciprocate a kind act of the first mover even if the first mover is much richer initially.

Rabin’s (1993) model of fairness only applies to two-person normal-form games with pure strategies. Others have developed more general versions of similar models (Dufwenberg and Kirchsteiger 2004, Falk and Ichino 2006). Such intentions-based models have been corroborated experimentally by, for example, Dhaene and Bouckaert (2010), who show that Dufwenberg and Kirchsteiger’s (2004) model is consistent with a large majority of people’s behavior in an experimental setting, and by Falk et al. (2008) and Blount (1995), who show that both intentions and outcomes are necessary to explain behavior.

There are a number of additional theoretical approaches that do not fit neatly into the three categories. For example, some models account for behavior that might be considered

---

4See also Geanakoplos et al. (1989) and Battigalli and Dufwenberg (2009).
reciprocal using bounded rationality (Roth and Erev 1995, Gale et al. 1995). Cox et al. (2007) present a model of reciprocity in which the weight placed on someone else’s material outcome is a function of one’s “emotional state,” which is in turn a function of the reciprocity motive, judged relative to a context-dependent neutral outcome. Sugden (1984) models reciprocity as an obligation: individuals feel obligated to contribute to a public good based on how much others are contributing. Yet other authors emphasize the role of evolution in supporting reciprocity as a motive, with a primary focus on the conditions under which reciprocity can be supported as an evolutionarily stable behavior (see Hoffman et al. (1998), Gintis (2000) and Bowles and Gintis (2011)).

2.3 The Missing Piece

One way to frame the limitation of the three broad classes of reciprocity models is that they assume that the opportunity to invest in others’ outcomes is always a welfare-enhancing expansion of the choice set. That is, the mere option to share with others cannot hurt the decision maker, but may increase utility.

As we reviewed above, the same limitation applied to the earlier models in ‘one-sided’ giving contexts. The key assumption used to be that individuals share because they like to share. More recently, however, the introduction of avoidance options into laboratory and field settings revealed that a majority of people share reluctantly. They share if asked, but prefer to avoid the sharing request. In response to these findings, the new wave of modeling approaches has focused on extrinsic factors, such as social pressure and social image.

Research on reciprocity has largely not yet incorporated these more recent approaches. The neglect of extrinsic factors could bias our understanding of reciprocity. Consider the following example. Suppose that, in a simple dictator game, 65 percent of dictators share. Now allow recipients to send a small gift to the dictators, prior to the dictator game, and suppose that the percentage of dictators who share increases to 80 percent. A simple model of intrinsically motivated reciprocity would attribute the decision of all 80 percent of sharers to welfare-enhancing motivations such as altruism. This interpretation might be wrong on two counts. First, some of the 65 percent who share in both the simple one-sided DG and in the two-sided reciprocity setting might be motivated by extrinsic factors, such as social pressure or social image concerns. These people would prefer to avoid the sharing situation, as found in Lazear et al. (2012). Second, the additional 15% who share only in the two-sided game may be changing their behavior due to changes in extrinsic motivations (such an increase in social pressure), rather than changes in intrinsic motivations (such as an increase in altruism induced by the recipient’s former kindness). These mis-inferences grow when we consider the choice of the quantity shared and of whether to avoid the sharing environment.

3 Experiment

The question of whether or not sharing in reciprocity environments is fully explained by intrinsic motives is an empirical question. One way to explore this question is to mirror the
approach from one-sided giving experiments and introduce avoidance options into reciprocity settings.

We report the results of a novel experiment intended to test how reciprocity affects avoidance. In particular, we apply the experimental approach of Lazear et al. (2012), which allows agents to avoid making a decision and keep all of their wealth, to a reciprocity setting. That is, we estimate the effect of a “sorting” opportunity after inducing reciprocity. Does a kind initial treatment reduce or eliminate avoidance, as implied by the existing theoretical motivations for reciprocal behavior? Or is the effect of kindness orthogonal to avoidance behavior?

### 3.1 Basic Idea

In a standard dictator game the decision maker has had no prior interaction with the (typically anonymous) recipient before making a decision. Hence, reciprocity is irrelevant, and the dictator’s choice only involves trade-offs between personal material payoffs, intrinsic motivation to share (e.g. altruism), and extrinsic motivations to share (e.g. social pressure). Previous experiments were able to identify external motives by introducing a sorting option into the dictator game. Lazear et al. (2012) gave dictators the option to avoid the decision to share and keep the endowment without the potential recipient learning about the sharing opportunity. They found that a large fraction of “sharers” decided to opt out, indicating the importance of external factors.

In this experiment, we use variants of a “double dictator game” (DDG) to test whether extrinsic determinants also affect sharing in a reciprocity environment. Here, dictator and recipient play a mini-dictator game prior to the main dictator game, with the dictator and recipient roles switched. The parties learn about the second stage (the main DG) only after the mini-game has been played. The purpose of the initial reversed mini-dictator game is to induce reciprocity as a motive for sharing in the second-stage main dictator game. Positive reciprocity is induced if the mini-dictator decides to share. Negative reciprocity is induced if the mini-dictator decides not to share. The standard dictator game (DG) is the neutral “no reciprocity” benchmark for comparison. That way, we avoid confounds with other social preferences, as cautioned by Cox (2004). We choose a “mini”-dictator game in the first stage in order to distinguish reciprocity from distributional preferences such as inequity aversion.

We cross these treatments with the option to avoid the sharing decision in the second stage and keep the endowment. Table 1 illustrates the 3x2 treatment design. The first dimension (columns) designates the reciprocity condition: the standard dictator game is the neutral setting, without any reciprocity involved; the second and third conditions use the decision of the mini-dictator to generate a positive or negative reciprocity environment. Following earlier evidence, we expect the greatest sharing under positive and the least under negative reciprocity. The second dimension (rows) designates the sorting condition, which changes the choice set of the dictator in the second stage. In the no-sorting conditions, the dictator is forced to choose an allocation, of which the recipient is then informed. In the sorting conditions, the dictator has the option to costlessly opt out of the game, thus receiving the full endowment but leaving the potential recipient uninformed about the game.
These six conditions will allow us to evaluate the role of intrinsic and extrinsic determinants of sharing. Under the existing models for reciprocity giving is driven by intrinsic motivations, and the sorting option should become irrelevant after reciprocity has been induced. That is, a positive reciprocity setting, in which the decision-maker has previously been treated kindly by the recipient, should amplify the weight placed on the desire to share sufficiently to render extrinsic factors such as social pressure irrelevant. Similarly, negative reciprocity induced by not sharing in the prior mini-game, should induce spite and a willingness to punish the other party, which renders extrinsic motivations such as social pressure irrelevant and makes people happy to share nothing even if they cannot opt out.

Thus, while we know that about half of the dictators who share in a simple dictator game prefer to sort out if possible, the existing models of reciprocity predict no such sorting behavior after reciprocity has been induced. The only people who might sort out are those who are indifferent between opting out and sharing nothing.

Alternatively, extrinsic motives are still at play after reciprocity has been induced. For example, it is possible that positive reciprocity increases the intrinsic motivation, but does not eliminate the extrinsic motivation. In this case, we would expect to see sorting out in the DDG after the mini-dictator has shared, but less than in the standard DG with sorting. We would also expect the average amount shared to increase. Another possibility is that extrinsic motives affect reciprocal behavior directly. For example, positive reciprocity could increase the image cost of being ungenerous. In this case the predictions regarding opt-out frequency and average amount shared are ambiguous. With both intrinsic and extrinsic determinants becoming stronger, both quantities could increase or decrease since sorting in is now more costly due to increased social pressure, and sorting out is more costly due to increased guilt. However, we do have a clear prediction for average sharing conditional on not opting out: those deciding to play the (main) dictator game will share more. We explore these possibilities in the context of a formal model further in section 5.

3.2 Experimental Design

To test for extrinsic determinants of reciprocity-induced sharing, we conducted a reciprocity variant of the dictator game, a “double dictator game” (DDG). The DDG consists of two dictator games played consecutively, with the role of dictator and recipient switched in the second stage. We compare the DDG to a standard dictator game (DG), and cross the design with sorting options. We employ a between-subjects design based on the design in Lazear et al. (2012) to compare games with and without sorting, holding constant the endowment. The design is similar to Dana et al. (2006) and to Broberg et al. (2007).

All experiments were conducted at UC Berkeley. Subjects received a participation fee of $5 and were informed that they might earn additional money. Half of the subjects were randomly assigned the role of recipient (and hence, in the DDG, the role of mini-dictator). The other half were assigned the role of dictators (and hence, in the DDG, of mini-recipients) and were moved to a separate room from the recipients, where they received their instructions.

DDG with and without sorting. The first-stage mini-dictators were told to divide $2
with a randomly-paired participant in the other room by circling one of two choices: keep $2 and give the paired participant $0; or keep $1 and give the paired participant $1. We offered only a binary choice (share $1 or $0) to generate clean reciprocity treatment assignments.

After these participants made their choices, participants in the other room were told about the first stage. After finding out how much of the $2 their paired participant shared, the matched partner played the second stage. In the variant without sorting, this was a $10 dictator game with the same person. Note that subjects did not know, initially, that their matched partners would subsequently become dictators over $10.\(^5\) At the end of the experiment, the experimenter described the game to the participants in the other room and showed each of them how much money they received. Thus, recipients were guaranteed to know about the second-stage game and about the amount dictators had decided to share with them.

In the variant with sorting, second-stage dictators first decided whether or not to “participate.” They received two envelopes, labeled “participate” and “don’t participate.” If they chose to participate, they opened the former envelope, which contained a sheet with the participant number of the paired recipient, and filled it out exactly as in the no sorting condition. If they chose not to participate, they opened the envelope marked “don’t participate” (which did not contain a matched participant number) and wrote their participant number on the sheet inside. In that case, they received $10 without the option to divide the money.

In this variant, participants in the other room only found out about the dictator game if their matched dictator had chosen to participate. After collecting the envelopes, the experimenter separated receivers matched with participating and non-participating dictators. Those matched with non-participating dictators received the $5 participation fee and the earnings from the mini-dictator game, and the experiment ended. Those paired with participating dictators completed the experiment as in the no sorting treatment, meaning they were informed about the dictator game and shown how much the dictator had shared.

A total of 192 pairs of subjects (54 without and 138 with sorting) participated in the DDG. In 89 cases (46 percent), the first mover shared the two dollars (26 cases [48 percent] without and 63 cases [46 percent] with sorting). This led to 89 dictators in the PR condition (63 with sorting and 26 without) and 103 in the NR condition (75 with sorting and 28 without).

**DG with and without sorting.** The benchmark for comparison of the reciprocity environment are baseline dictator game (DG) conditions with and without avoidance options. The procedures for the DG variants closely mirror those of the DDG variants, just without the first stage\(^6\). Details are described thoroughly in Lazear et al. (2012). Overall, 182

---

\(^5\) Initial voluntary sharing can thus be interpreted as an act of kindness, rather than an attempt to induce reciprocal behavior, which avoids concerns about interpreting the dictators’ reactions as reciprocity.

\(^6\) One difference to the DDG is that the instructions emphasized the $10 dictator game (including the decision not to participate in case of the sorting treatment) was the last decision that anybody in either room would make in this experiment. This did not seem necessary in the standard DG, consistent with how the DG is conducted in prior research.
student subjects participated in the DG treatments, including 45 dictators in the no-sorting condition and 46 in the sorting condition.

### 3.3 Experimental Results

Before we turn to the main results on the effect of sorting in reciprocity environments, we briefly discuss the evidence of reciprocity in the conditions without sorting.

Figure 2 presents the average amounts shared, without sorting (left, dark columns), with sorting and treating those who opt out and thus share nothing as having shared $0 (middle, striped columns), and conditional on sorting in (right, grey columns). The three sets of bars show average levels of sharing in the three reciprocity conditions: the baseline DG (no reciprocity), positive (PR), and negative (NR) reciprocity. Focusing on the conditions without sorting (left columns in each set of bars), we see that the average amount shared by dictators in the standard (no-sorting) DG treatment is $2.00 in our experiment. In the (no-sorting) DDG, the average amount shared by dictators increases to $2.39 in the PR treatment, and it decreases to $0.70 in the NR treatment. Table 2, Column 1, shows that only the negative-reciprocity effect is significant. In other words, there is a significant negative-reciprocity effect (−0.130) and an insignificant positive-reciprocity effect (0.039) which is consistent with previous experimental evidence, e.g., weak positive reciprocity but strong “concern withdrawal” in Charness and Rabin (2002).

Note, though, that the above averages are based on the second-stage endowment of $10 and, hence, assume “narrow bracketing” in the second-stage dictator game. An alternative measure of reciprocity adds the amount of $2 from the mini-DG back to the analysis (see Cox (2004)), in which case the effect of positive reciprocity is marginally significant and the effect of negative reciprocity is insignificant. The subsequent analysis on the effects of sorting is unchanged when $12 is used as the relevant endowment.

The Effect of Sorting on Sharing. We now turn to the question of what impact sorting has on sharing in the baseline DG and each of the reciprocity environments. As mentioned above, the average amount shared in the DG treatment without sorting is $2.00, which is comparable to findings in previous experiments. Most subjects share a positive amount (64 percent). However, the introduction of sorting strongly decreases the average amount shared, to $1.21, as shown in the left set of bars in Figure 2. This decrease is statistically significant in a non-parametric rank-sum test ($z = 2.34, p = 0.02$). The sorting

---

7 Standard errors are robust to heteroskedasticity and adjusted for small-sample bias, using the residual-variance estimator HC3, which approximates a jackknife estimator (MacKinnon and White 1985). If we cluster by session, standard errors in this and in all other estimations are very similar and typically slightly smaller, though unlikely to be reliable given the few clusters.

8 Under this alternative approach, recipients end up with an average amount of $1+$2.39 = $3.39 out of $12 (28.3%) after sharing $1, and with an average payoff of $2 + $0.70 = $2.70 (22.5%) after sharing zero, compared to $2 out of $10 (20%) in the single dictator game. In this case, positive reciprocity induces a marginally significant increase in giving ($t$-statistic = 1.88, $p$-value = 0.06), and negative reciprocity does not have a significant effect. The lack of a significant negative-reciprocity effect reflects censoring at $2: Dictators cannot reduce the amount obtained by recipients below $2 if those kept the initial $2.
opportunity also decreases the frequency of sharing dramatically, to only 39% sharing a positive amount.

How does sorting affect sharing after positive or negative reciprocity has been induced? In Figure 2, we see that positive reciprocity increases giving to an average level of $2.39 without sorting (left bar in middle set). But, once again, sorting causes a large drop in average amounts shared, to $1.71 (middle bar in middle set). In a negative reciprocity environment without sorting, sharing plummets to an average of $.70, and the option to sort out further reduces sharing to $.31 (left and middle bars in right set).

Figure 1 provides more details and shows the distribution of amounts shared in each condition. (We display the frequencies of subjects who opt out separately at the left.) We observe a sharp shift of the distribution to the left when sorting becomes possible, regardless of the reciprocity conditions. Hence, both the simple comparison of means and the distributional evidence suggest that sorting has a large impact on sharing even after reciprocity has been induced, inconsistent with a solely intrinsic motivation for reciprocal behavior.

Finally, Table 2 confirms the statistical significance of these findings, both in a linear regression (column 2) and in a tobit estimation (column 4). As before, standard errors are robust to heteroskedasticity and, in the linear regression, adjusted for small-sample bias. Under both estimation procedures, sorting significantly reduces sharing.

We conclude that sorting retains its impact even after we induced either positive or negative reciprocity. The option to sort significantly decreases sharing in both reciprocity conditions. In other words, givers who respond to a previous kind or unkind act are affected by the option to avoid the opportunity to give. This evidence suggest that the dominant approach to sharing under reciprocity, which relies on intrinsic factors, is incomplete. Extrinsic factors affect individuals’ giving in reciprocity environments as well, inconsistent with the prevalent modeling approach.

But how strong are extrinsic factors after the inducement of positive or negative reciprocity? Are they identical to the degree found in neutral settings, or does reciprocity lead to differential sorting effects? The simple comparison of means suggested that the effect is smaller under positive reciprocity and larger under negative reciprocity than in the neutral DG setting, though the differences did not seem large. Looking back at Figure 2, we see that the average amount shared decreases by 40 percent, from $2.00 to $1.21, in the DG condition; by 29 percent, from $2.39 to $1.71, in the PR condition; and by 56 percent, from $0.70 to $0.31, in the NR condition.

The visual impression holds up to statistical testing. As Columns 3 and 5 of Table 2 reveal, the significant decrease due to sorting does not differ significantly, from the DG condition, in either the NR or PR conditions. The same picture emerges if we consider the frequency of sharing, i.e., the fraction of subjects who share any positive amount. The probit regression in the final column of Table 2 shows that 25 percent of sharers opt out,

---

9 Following MacKinnon and White (1985), we use the residual-variance estimator HC3 in the linear regressions, which approximates a jackknife estimator. In the tobit model, we perform a jackknife estimation, which produces slightly more conservative standard errors than the robust variance estimator.
but the interactions of Sorting with either reciprocity condition, PR or NR, are statistically insignificant. In other words, the impact of sorting on average amounts shared is large and significant, and approximately invariant to the reciprocity setting. We will investigate these reduced-form findings in more detail in our structural estimation in section 5.2 where we provide quantification of the role of intrinsic and extrinsic factors.

**The Effect of Reciprocity on Sorting.** The above analysis of amounts shared revealed that reciprocal behavior appears to be affected by extrinsic factors. Otherwise, giving would have been unaffected by the opportunity to avoid the giving request. We can go one step further and analyze to what extent the reduction in giving comes via the channel of opting out. In other words, compared to the simple DG, how does reciprocity impact the choice to opt out?

The leftmost bars in each of the three graphs in Figure 1 reveal that the sorting option is used by a significant fraction of subjects in all treatments. In the DG condition, 50 percent of all subjects choose to opt out. In the PR condition, many still sort out, but fewer than in the DG baseline (32 percent). The NR condition incites more sorting behavior than the PR condition, but only slightly more than the baseline DG (59 percent). The differences between sorting under reciprocity and sorting in the neutral DG setting are either insignificant or only marginally significant. If we regress a dummy for “opting out” on a dummy for being in the PR treatment and a dummy for the NR treatment, as shown in column 1 of Table 3, the intercept of .500 is highly significant, while the PR coefficient of -.183 is only marginally significant, and the NR coefficient of .087 is insignificant. The results are similar if we use a probit model, though the PR coefficient becomes more significant (marginal effect of -.186 with s.e. .094), as shown in column 2.

The sorting evidence speaks to the presence of both intrinsic and extrinsic factors affecting reciprocal sharing decisions. If sharing was fully explained by intrinsic determinants then subjects should not make use of the option to sort out. If anything subjects who do not share would be indifferent. Instead, we find that a significant fraction of subjects use the sorting option even after reciprocity has been induced. At the same time, we do see that the use of the sorting option decreases after positive reciprocity has been induced, suggesting that a kind initial treatment strengthens the intrinsic motivation. But, by the same logic, negative reciprocity should make people intrinsically less willing to share, and therefore also less likely to avoid the sharing environment. In other words, we would expect that non-sharers are not afraid to sort in and reveal their non-sharing decision (anonymously) to a partner who did not share himself. The small and insignificant increase in opting out in the NR treatment, relative to the DG, along with the large observed drop in actual sharing discussed above, hints at problems with this interpretation.

To be clear, we do observe variation in “spiteful non-sharing,” i.e., in the fraction of subjects who sort in but share nothing. As can be seen in Figure 1, only a small number of subjects sort in and share zero in the DG (11 percent). Positive reciprocity reduces this rate to 3 percent (a 72 percent reduction). Meanwhile, in the NR condition, 20 percent of dictators sort in to share nothing (a 46 percent increase over DG).

As columns (3) and (4) of Table 3 reveal, however, the differences are not significant in an
OLS regression, and only the reduction after kind treatment (PR) is marginally significant in a probit estimation. However, the change in the rate of spiteful non-sharers between NR and PR is highly significant under both econometric models. Given the magnitude of the effect and the strength of the PR/NR comparison, and the fact that the effect is likely underestimated due to censoring at zero, we can infer that reciprocity has an impact on spiteful non-sharing.

Whether or not such behavior truly reflects spite, as suggested in the reciprocity literature, might be debatable. But it does indicate that non-sharers do not feel significant extrinsic pressures. It also suggests, though, that sharing and not-sharing is not fully explained by intrinsic motives. An intrinsically motivated non-sharer would not necessarily have the desire to let the recipient know about not sharing.

The Distribution of Gifts. As a final avenue through which we can examine the impact of sorting we zoom into different areas of the distribution of shared amounts. As we saw already in Figure 1, the introduction of a sorting option significantly changes the distributions of positive gifts in all three conditions. The graphical evidence suggested a strong shift to the left in all three graphs.

But how does the sorting option affect different regions of the distribution? To investigate this question, we categorize dictators into three groups: zero-sharers (including those who opt out), low sharers (who share but less than 25 percent of the endowment), and high sharers (who share at least 25 percent). Across all conditions, 52 percent are zero-sharers, 26 percent are low sharers, and 22 percent are high sharers.

We then regress indicator variables for each of those groups on an indicator for the sorting option, separately for all three groups. As shown in Table 4, the sorting option significantly increases the fraction of subjects who share nothing, by 20.9 percent (column 1). This is consistent with the strong sorting effect across treatments. We also find that the sorting option has no significant effect on the fraction of low sharers (column 3). This reflects the switch of low sharers to zero sharing, on the one hand, and the switch of high sharers to low sharing, on the other hand. Finally, the sorting option significantly reduces the fraction of high sharers, by 16.0 percent (column 5).

We also find that the effect of the sorting option in the three areas of the distribution does not vary significantly across the neutral, positive-reciprocity, and negative-reciprocity settings. Columns (2), (4), and (6) show that the interaction of the Sorting indicator and dummies for the positive-reciprocity and the negative-reciprocity settings are statically insignificant or at most (in one case) marginally significant. In other words, the effect of the sorting option in different areas of the distribution appears to be surprisingly similar across different treatments, which is hard to reconcile with the leading theories of reciprocal behavior.

---

10 Results do not substantially change if slightly different category definitions are used.
4 Discussion

What are the implications of these results for our understanding and theories of reciprocity? Existing theories have to wrestle with two key findings:

1. In all three conditions, the sorting option has a significant impact on the distribution of gifts and the average gift, and the impact is approximately invariant to reciprocity.

2. Moving from DG to PR, we observe less opting out and less spiteful non-sharing. Moving from DG to NR, we observe more opting out and more spiteful non-sharing. The increase in spiteful non-sharing under NR is economically large, and the difference to the PR treatment is highly significant.

We briefly recap why existing models fail to predict those findings. To preview the gist of the discussion, all models predict that the sorting option has zero impact on the average or the distribution of gifts, since the only people who sort out are those who are indifferent between sorting out and sharing zero. This contradicts the first point. Additionally, without richer assumptions regarding indifference, the fraction who sort out and the fraction who sort in and share zero (spiteful non-sharers) are predicted to increase and decrease in proportion to each other, invariant to the reciprocity setting. This contradicts the second point. Additionally, the results comparing the impact of sorting across reciprocity settings are not spoken to by these models, due to their common prediction of zero impact of sorting in all cases. We propose several alternative explanations that might fill this gap.

In discussing the predictions of the outcome-, type-, and intentions-based reciprocity models from Section 2.2 we will need to make an additional assumption to solve the cases of indifference. All models imply that non-sharers are indifferent between sorting out of the dictator game and sorting in (and keeping all of the money). We assume that dictators choose to sort out with some fixed probability that is not context-dependent. This assumption allows, in particular, for “sorting out as default” and “randomize” as possible heuristics.\footnote{More complicated rules may improve compatibility of theory and data, such as allowing the probability of sorting out to depend on type or context, but would require the model to formulate these dependencies more explicitly, so we do not discuss these possibilities in this section.}

First consider outcome-based models, such as Fehr and Schmidt (1999). In the DG, types who put sufficiently high weight on fairness or others’ consumption will share a positive portion of the $10. Everyone else will be indifferent between sorting out or sorting in but sharing nothing. In the DDG, however, the initial distribution of an additional $2 might induce the dictator\footnote{In this section, labels of players always refer to their role in the primary (second-stage) dictator game. For example, the decision-maker in the mini-DG is referred to as the recipient. Throughout, male pronouns are used for dictators, and female for recipients.} to share more or less to achieve his most preferred distribution of wealth. In the negative reciprocity condition, the recipient has a head start of $2. Hence, a higher threshold for weight on fairness is required for the dictator to share a positive amount, and positive amounts shared will be smaller. In the positive reciprocity condition, dictators and recipients both have a head start of $1. Depending on the exact model formulation,
this could cause the threshold level of fairness for positive sharing to go up (if dictators no longer feel the need to ensure a small minimum payment for the recipient) or down (if the marginal utility of 11th dollars is particularly small) or stay the same (if dictators are merely inequity-averse), relative to the DG treatment. However, the mere introduction of a sorting option has no impact on the distribution of positive gifts. Instead, any change in sharing behavior between reciprocity conditions is driven by a shift in the distribution of choices upwards or downwards, with the portion of the distribution censored at 0.

These models therefore successfully predict that the distribution of giving shifts downward in the negative reciprocity case, compared to the baseline DG. They do not necessarily capture the increase in sorting in and in average gifts in the positive reciprocity conditions. With our additional assumption that non-sharers who are indifferent between sorting options choose to sort out some with probability $p$, they also predict that any change in actual sharing should be accompanied by a change in sorting choices. This is indeed the case in the positive reciprocity game (more people sort in to share a positive amount). But it is not the cases in the negative reciprocity treatment. Here, we find an increase in spiteful non-sharing without a proportional increase in sorting out. In fact, $p$ would need to be larger than in the DG to explain the NR results, but would need to be somewhat smaller than in the DG to explain the PR results.

Additionally, note that outcome-based models of reciprocity trivially predict that the diff-in-diff effects of a change in the impact of the sorting option across reciprocity settings is zero, since they predict zero impact of a sorting option in any context. We in fact find a small (insignificant) diff-in-diff in the level of average gifts, but it results from large, statistically similarly sized impacts across settings (see Figure 2 or Table 2).

Next consider type-based models, such as Levine (1998). In these models, the dictator’s altruism towards the recipient depends on his information about the recipient’s type. In the standard game, then, since the dictator has no information about his partner, he simply chooses his optimal wealth allocation given his knowledge about the population distribution of types. In the reciprocity conditions, however, the dictator can infer something about his partner based on whether he received $1 or $0. After receiving $1 in the PR condition, then, the dictator has strictly more favorable beliefs about the recipient than in the DG condition, and should place a higher weight on her consumption. In the NR condition, conversely, the dictator places strictly less weight on the recipient. Hence, type-based models can explain the increase in sorting in and giving that we observe in the PR conditions. As in the outcome-based models, however, this directional prediction is still not unambiguous: dictators may be less likely to share in the PR than in the DG condition, despite putting a higher weight on altruism, because the recipient is already starting out with $1. Moreover, as in the outcome-based models, any changes in giving are driven by a shifting distribution of ideal allocations, and are not affected by a sorting option. Hence, the results about differential effects of the sorting option along with the rates of spiteful non-sharing cannot be explained by type-based reciprocity.

---

13 Since the recipient was not informed about the main dictator game prior to making her choice in the mini game, this choice cannot be interpreted as a strategic signal.
Lastly, consider intentions-based models. Since the recipient does not know about the later dictator game when she makes her mini-DG choice, we can analyze behavior as two normal form games and apply the version of the model in Rabin (1993) (even though our game is technically an extensive form game). Also, note that, in our context, both players know their partner’s actions (within the game they believe they are playing) so that there is no concern about distinguishing beliefs from actions.

In the DG, the recipient has taken no action and is thus neutral. The dictator should therefore maximize his material outcome. Likewise, in the initial mini-dictator game in the DDG, the recipient believes that the dictator cannot respond and is thus neutral, and should also maximize her material outcome by keeping the $2. The recipient’s “kindness” in the mini-dictator game measures the impact of her choice on the dictator’s outcome relative to the neutral average. Hence, it is negative if she shares nothing and positive if she shares $1. In the negative case, the dictator simply maximizes his material outcome as there is no avenue for costly punishment of the recipient. In the positive case, however, any dictators with a high enough weight on reciprocating intentions shares a positive amount. (Note that since players do not have utility over their partners’ outcomes directly, but only over each others’ reciprocal intentions, this model makes unambiguous predictions in the PR game.) However, this model predicts, once again, that the sorting option should have no impact on the distribution of positive gifts in any condition. Hence, the other conflicts with the data are not addressed by intentions-based reciprocity any better than with outcome- or type-based models.

So what can explain those findings? The sorting effect in the simple DG shows us where to look for an explanation for the sorting effect in the DDG as well. Between the standard dictator game and the same game with a sorting option, the only change is that the information acquired by the recipient becomes manipulable by the dictator. If the dictator sorts in, the recipient will know how her monetary outcome was determined, but if he sorts out, she will not. Anyone who modifies his behavior when given a sorting option must do so in response to this additional choice over information. Since this response to observability is not included in other models of reciprocity, those models are necessarily incomplete.

As discussed in the introduction, we argue that the “missing piece” in existing models is extrinsic motivation. Individuals may share not because they care about the other person’s payoff, but because they feel obliged or pressured to and would rather avoid being in this situation. There is a variety of terminology describing extrinsic determinants, including social image, social pressure, social signaling, audience effects, prestige, shame, guilt, and reputation. We will sometimes refer to “social image” or “social pressure,” but our analysis does not pin down the exact form of extrinsic motivation. Rather, it illustrates that a model of reciprocity that incorporates extrinsic factors can predict and provide an interpretation for our results.

14 A more realistic version of the material outcome utility might incorporate some other form of social preferences such as inequity aversion or pure altruism; but to avoid misattributing implications of these other models to intentions-based preferences, we accept for now the strict material interpretation.
Before we turn to the model and structural estimation in the next section, we briefly provide the intuition for our approach. Suppose that a dictator trades off monetary payments, intrinsic factors such as fairness or distributional preferences, and extrinsic factors such as social image and social pressure in deciding whether to share. In the standard DG, a dictator who shares nothing would prefer to sort out if a sorting option becomes available, as this leads to identical outcomes but a better social image. A dictator who shares a positive amount in the standard DG might also prefer to sort out if possible. This would be the case, for example, if the material benefit from keeping the additional money and the benefit from not feeling social pressure to give more than intrinsically desired outweigh the costs of possibly giving less than desired and not obtaining social image benefits. In either case, extrinsic determinants are needed to predict sorting out. Additionally, sorting could impact the distribution of gifts. For example, if generous givers in the standard DG are mostly individuals who are particularly susceptible to social pressure, they will make use of the sorting option if available and the distribution of amounts given will change significantly if the option to sort becomes available.

In the positive reciprocity scenario of the DDG, extrinsic factors might become less important as positive reciprocity strengthens the internal (altruistic) motivations for giving. Alternatively, positive reciprocity might strengthen extrinsic motives and make it less acceptable to share little. If this manifests as a larger image reward for sharing, people might share more. If it manifests instead as a larger image punishment for imperfect sharing, it might drive more people to sort out. In the negative reciprocity scenario, less sharing might reflect not only lower intrinsic motives to share, but also lower extrinsic motives. For example, the pressure to share may becomes lower and a dictator who would share a positive amount in the no-sorting setting might now believe it to be fair to share nothing. Alternatively, the intrinsic willingness to share may become lower and the dictator might share nothing. However, only extrinsic factors such as social image concerns predict an increase in sorting in to share zero, namely as a boost to social image resulting from punishing unfair peers or as a relief from external pressures to share.

Hence, a model that allows extrinsic determinants of giving is able to explain the reduced form findings, including changes in donations and sorting-in rates that are not positively correlated. The structural estimations will provide a clearer picture of the relative strength of the underlying motivates and their interactions with reciprocity than we can glean from the reduced form analysis.

5 Model and Estimation of Extrinsic Motives

We now present a simple model of reciprocity that incorporates both intrinsic and extrinsic determinants of sharing. Despite its simplicity, this model explains the key aspects of the experimental results much better. Many models of social image and other extrinsic determinants of reciprocity are possible, and further experimentation is needed to pin down the exact form of preferences. Nevertheless, following the logic above, even a simplistic model of image-based reciprocity conveys the underlying mechanism. We confirm this using a
structural estimation, described in the next section.

5.1 Model

The outcome-, type-, and intentions-based models of reciprocity described above can be thought of as describing intrinsic determinants of sharing whose weights vary with the reciprocity environment. That is, the dictator keeps an amount $x \in [0, 10]$ such that

$$U_r(x) = x - \alpha_r G_r(x),$$

where $\alpha_r G_r$ captures intrinsic motives to sharing as disutility from keeping too much. $G$ is an increasing function of $x$, which may vary depending on the reciprocity environment $r \in \{DG, NR, PR\}$, and $\alpha_r \in \mathbb{R}$ is the weight assigned to sharing the “right amount” and may also vary with the reciprocity environment. As before, we will assume that non-sharers, who are indifferent between sorting out and sorting in (and keeping all of the money) sort out with some fixed probability that is not context-dependent.

The simplest way to incorporate extrinsic factors into this framework is to add a parallel weight $\beta_r \in \mathbb{R}$, applied to an increasing function $H$, which kicks in only if the dictator sorts in, i.e., if his actions are observable to the recipient:

$$U_r(x) = x - \alpha_r G_r(x) - \beta_r \mathbb{1}(\text{sort in}) H_r(x)$$

To minimize the degrees of freedom and be able to identify the model, we impose the following specification:

$$U_r(x) = x - (\alpha_r + \beta_r \mathbb{1}(\text{sort in}))(x - 5)^2$$

A few details deserve comment. First, the quadratic loss function around the (narrowly-bracketed) 50-50 split implies that people want to be generous but not too generous. This functional form predicts that no one will give more than $5, and indeed almost no one does. It also has the benefit of predicting the second mode of giving, at the 50-50 split: anyone above a certain threshold of intrinsic motivation will share exactly $5. We will verify in the robustness checks in Section 5.2 that the specification is not critical to our findings.

Second, the fact that the loss function is symmetric around $5 in all reciprocity treatments means we are assuming narrow bracketing, rather than incorporating the initial $2 mini-DG into the utility function (see the discussion at the beginning of section 3.3). However, a global-bracketing 50-50 split would again prescribe sharing 5 in the PR condition, and sharing 4 in the NR condition. This appears to be irrelevant to our findings. We also note that we continue to observe a few people sharing exactly $5 in the NR condition, and no comparable bump in the distribution at $4 (see Figure 1c).

Third, the fact that the extrinsic factor is framed as a loss ($-\beta_r \mathbb{1}(\text{sort in}) H_r(x)$) implies that a giver cannot gain utility (as long as $\beta$ is positive). At best, he obtains no loss in

---

15 This is also the view supported by, e.g., Andreoni and Bernheim (2009). The debate over whether over-generosity induces guilt and a bad social image is still active (see Krupka and Weber (2013)), but is not relevant to our estimations.
utility, namely when sharing the “fair” amount of $5. For example, in our specification in (1), sharing less than $5 can never lead to an increase in utility. And, vice versa, if $\beta < 0$, i.e., if there is a reward for punishing the other person, then sharing any amount leads to an increase in utility. This is almost certainly not true. It seems plausible that being “sufficiently fair” or giving “sufficient punishment” would suffice to break even. The arguments apply also to the intrinsic motivation for sharing. However, this simplification merely limits our ability to estimate the absolute utility impact of presenting someone with an opportunity to share. (Our treatments are not designed to identify the cutoff for giving to be welfare increasing.) We relax this assumption in several robustness checks and verify that this assumption is not critical for our results (see Section 5.2.)

Fourth, treating $\alpha$ and $\beta$ additively implies that internal and external determinants of sharing operate in parallel, with one exception: an agent may feel spiteful despite societal pressure to share, or may feel compelled to share despite pressure to punish. That is, the two parameters may have opposite sign.

The sharing implications of this model are straightforward. In the absence of social pressure or other extrinsic factor, dictators would like to keep the amount $x^* = 5 + 1/(2(\alpha r))$ (rounded to the nearest available discrete choice)\(^{16}\) But if they sort in, they will face social pressure and hence keep the social pressure-adjusted amount $x^s = 5 + 1/(2(\alpha r + \beta r))$\(^{17}\) If the utility obtained with this adjusted choice is smaller than the utility when selfishly opting out and sharing zero, they will sort out if possible. Higher intrinsic motivation to share increases the cost of sorting out, and higher social pressure increases the cost of sharing moderately.

Figure 4 illustrates how the two type parameters, $\alpha$ and $\beta$, determine choices when sorting is allowed, for a given reciprocity environment. When $\beta < 0$, no one chooses to opt out. That is, if selfishness induces a good social image, there is no reason to opt out rather than visibly sharing nothing. Furthermore, more negative levels of $\beta$ must be met by increasingly high levels of $\alpha$ in order for people to be persuaded to share anything at all. When $\beta > 0$, instead, higher levels of $\alpha$ still lead to more generosity among those who sort in, and higher levels of $\beta$ force relatively ungenerous people to share more generously. But now people who want to give small amounts rather sort out and feel a bit guilty for being selfish than sort in and be seen as greedy.

The figure shows how the availability of a sorting option influences the average level of giving and the distribution of gifts (among those who share) in different ways. At low levels of $\beta$, introducing a sorting option will most affect the lowest givers, so the impact on average generosity should be small. When $\beta$ is higher, many generous givers (with low $\alpha$’s) will also opt out. Conversely, at low levels of $\alpha$, a sorting option will affect givers, and the opting out of the most generous givers (with high $\beta$) will have a large impact, while givers higher levels of $\alpha$ are more immune to the sorting option overall.

The figure also illustrates why existing models of reciprocity fail to explain the aspects

\(^{16}\)Note that $x^*$ is a maximum if and only if $\alpha r > 0$. If $\alpha r < 0$ then $x^*$ is a minimum. Therefore, in this case, we would have a corner solution.

\(^{17}\)Note that $x^s$ is a maximum if and only if $\alpha r + \beta r > 0$. Otherwise, we will have a corner solution. This becomes pertinent in the negative reciprocity case, in which we estimate that $\alpha r + \beta r < 0$. 

22
of our data discussed in Section 3.3. Looking at Figure 4 along the $\beta = 0$ axis, the only individuals who care to sort out are those who are also willing to sort in and share nothing. Hence, the systematic decrease after PR and increase after NR is not predicted. Additionally, a mere shift in intrinsic motivation cannot explain a change in the distribution of gifts since, again, the only people willing to sort out are those who would share nothing in any case.

Finally, another way to look at Figure 4 is to consider the possibility that the distributions of $\beta_{DG}$, $\beta_{PR}$, and $\beta_{NR}$ are the same, while the distributions of the $\alpha_r$’s may vary, such as $\alpha_{NR} < \alpha_{DG} < \alpha_{PR}$. The figure shows that, as the distribution of alpha moves to the right, people share more and opt out less. In other words, the impact of sorting on the unconditional average level of sharing and rate of sorting should be smaller when the recipient has acted more kindly.

Table 5 summarizes the predictions of three possible variants of the model in terms of the patterns we observed in the data. All three variants allow for intrinsic motivations, such as altruism or fairness, and assume that their strength is increasing in the prior kindness of the recipient. (For the purpose of the table, we consider reciprocity $r$ to be a continuous variable, where a more positive value means a more positive reciprocity context, i.e., kinder prior treatment.) The first variant, shown in the first row, does not allow for extrinsic motivation ($\beta = 0$). This variant captures the predictions of the outcome-, type-, and intentions-based models discussed earlier. The second variant, shown in the second row, allows for extrinsic motivation, but assumes that it is unaffected by prior kind treatment. The third variant, shown in the third row, also allows extrinsic motivation to increase in the degree of prior kindness. The last row shows the data.

While all models are consistent with reciprocity increasing the average amount given and decreasing the frequency of opting out, a model without extrinsic motives cannot explain why the use of the sorting option and the distribution of gifts vary systematically with reciprocity, as indicated in columns 4) and 5). (Note that, as discussed, the data is not entirely clear on whether this prediction holds.) Moreover, column 3) shows that the increase in spiteful non-sharers in the negative-reciprocity setting is the main reason why any extrinsic factor that is constant across reciprocity treatments seems insufficient.

The table also clarifies that the reduced form results are inconclusive about the relationship between the reciprocity environment and the relative magnitudes of $\alpha$ and $\beta$. (See the “possible” entries in row 3, and “maybe” in row 4.) Our structural estimation will paint a clearer picture of the impact of reciprocity on intrinsic and extrinsic motivations for sharing. The variable rates of opting out, across treatments, and the change in the distribution of shared amounts conditional on sorting in are the main sources of variation that allow us to identify $\alpha$ and $\beta$ from the data.

### 5.2 Estimation

We estimate the parameters $\mu_{\alpha_r}$, $\sigma_{\alpha_r}$, $\mu_\beta$, and $\sigma_\beta$ for each $r \in \{DG, NR, PR\}$, using a minimum distance estimator given by $(m(\theta) - M)^\prime W (m(\theta) - M)$, where $M$ is the vector of true moments given by the data, $m(\theta)$ is the vector of moments predicted by the theory at vector of parameters $\theta$, and $W$ is the weighting matrix. For $W$, we use the diagonal of the
Inverse of the variance-covariance matrix. For the vector of moments, we break down the choices of giving into bins: exactly 0, from 25 cents to $2.50, from $2.75 to $4.75, exactly $5, and more than $5. In the sorting conditions, an additional moment specifies the fraction who sort out. Altogether, we have 11 moments in each reciprocity environment, or 33 total. In the baseline estimations, we also assume that $\alpha_r$ is normally distributed according to $N(\mu_{\alpha_r}, \sigma_{\alpha_r})$ and $\beta_r$ is similarly distributed according to $N(\mu_{\beta_r}, \sigma_{\beta_r})$. We will vary both the bin sizes and the distributional assumptions below.

The theoretical moments are simulated in Matlab using adaptive Simpson quadrature for numerical integration, implemented as the quad routine. An individual $i$ with type parameters $\alpha_i$ and $\beta_i$ in a particular reciprocity environment will share $x^s = 5 + 1/(2(\alpha_i + \beta_i))$ (or the closest element of the discrete choice set) if he cannot opt out; he will sort in and share $x^s$ even if he can opt out if $U(x^s) > 10 - 25\alpha$ (and otherwise will sort out). This threshold allows us to simply integrate over the distribution of types within the respective intervals to calculate the total fraction that fall within each choice category.

We determine the vector of parameters $\hat{\theta}$ that minimizes the distance estimator using fmincon, which is Matlab’s implementation of Powell’s (1983) sequential quadratic programming algorithm. We impose the constraints that $\sigma_{\alpha_r}$ and $\sigma_{\beta_r}$ are positive and at most 20, and that $\mu_{\alpha_r}$ and $\mu_{\beta_r}$ are between -20 and 20. To make sure to find the global minimum, rather than a local minimum, we choose starting points randomly from a uniform distribution on the allowable ranges and run fmincon on 5 to 120 of these starting point vectors, depending on the model specification, until the best estimates were clearly in concurrence. The best estimate is typically found in at least half of the runs.

The minimum distance estimate is typically asymptotically normal, with a variance of $(\hat{G}'\hat{W}\hat{G})^{-1}(\hat{G}'\hat{W}\hat{\Lambda}\hat{W}\hat{G})^{-1}/N$, where $\hat{\Lambda} \equiv \text{Var}(m(\hat{\theta}))$ and $\hat{G} \equiv \frac{1}{N} \sum_{i=1}^{N} \nabla_{\theta}m_i(\hat{\theta})$ (Wooldridge 2002). We calculate $\nabla_{\theta}m_i(\hat{\theta})$ numerically using an adaptive finite difference algorithm.

Column 1 of Table 6 shows the results. We estimate $\mu_{\alpha}$ in the dictator game to be significantly negative. The estimate indicates that a majority of people does not like to share at all. While our model is simplistic, this finding generally accords with our and others’ experimental findings that altruism towards strangers is widespread but far from universal. This baseline estimate implies that 31% of people would share a positive amount in a completely anonymous, dictator game with zero social pressure, which is not far from the observed rates of giving in experiments that have attempted to create this kind of setting (Koch and Normann (2008), for example).

What is striking is the magnitude of $\beta$ in the baseline DG. With $\mu_{\beta} = 2.56$ and $\sigma_{\beta} = 3.159$, fully 73% of people feel pressure to share, dwarfing the fraction who truly want to share, and the magnitudes of these estimates are significantly greater than zero. This adds to the mounting evidence that non-reciprocal (one-sided) giving is perhaps more driven by external

---

18 The same pattern of results emerges if using the identity matrix as the weighting matrix instead, but we omit these robustness checks for brevity.

19 These bounds are imposed in order to speed convergence. We verify that the global minimum lies within that range with runs using much larger bounds.
considerations such as social image than by internal social preferences.

The picture changes when reciprocity is involved. When the recipient has previously behaved selfishly (NR treatment), internal altruism parameters plummet to the point where only 16% of second-stage dictators feel any drive to share. At the same time, the weight of extrinsic factors seems to drop. The decrease might reflect reduced pressure to share, or increased pressure (or at least a license) to punish the recipient by answering selfishness with selfishness. Extrinsic determinants such as social pressure and social norms are usually thought of as attributes of a situation that apply equally to everyone; but since individuals might have different beliefs about how their actions are judged in this case, there is room for a mixed interpretation in which social pressure is reduced for some and inverted for others.

In the positive reciprocity environment, however, the weight of extrinsic factors returns to a very similar levels as in the baseline game. At the same time, the distribution of altruism moves upwards and a much smaller amount of the mass falls below zero. A majority, 52%, of people are predicted to return a favor to a kind partner, even if the partner would never find out.

Columns 2 through 7 show that these findings are robust to a number of important robustness checks. In Column 2 we show the results of an estimation using a larger set of 39 moments. We break the choice set into a finer categorization, namely, exactly $0, 25$ cents to $1.50, 1.75$ to $3, 3.25$ to $4.75, 5$ exactly, or more than $5$. The pattern of results is very similar, with $\mu_\alpha$ generally being slightly more negative and $\mu_\beta$ being more positive.

We also estimate specifications were we alter the assumption that $\alpha$ and $\beta$ are normally distributed, and assume a uniform distribution, with $\alpha_i \in [\mu_\alpha - \sigma_\alpha \sqrt{3}, \mu_\alpha + \sigma_\alpha \sqrt{3}]$ and $\beta_i \in [\mu_\beta - \sigma_\beta \sqrt{3}, \mu_\beta + \sigma_\beta \sqrt{3}]$, within each reciprocity environment. These are shown in columns 3 and 4 for the model with 33 and with 39 moments respectively.

As another robustness check, we also alter the model specification

$$U_r(x) = x + (\alpha_r + \beta_r \mathbb{1}(\text{sort in}))(f - (x - 5)^2).$$

In the baseline specification, $f = 0$. The alternate specification allows for $f > 0$, which means that extrinsic factors can increase the giver’s utility, e.g., in the form of pride, when giving at least a threshold amount. We estimate this model with $f = 5$ (column 5) and $f = 15$ (column 6), which implies that giving at least $2.76$ or $1.13$, respectively, is sufficient to feel pride or to be viewed favorably by your peers. The results are again very similar for the specification with $f = 5$. Mean intrinsic motivation is lower than in the baseline specification, very similar to the estimation with 39 moments in column 2, though the standard deviation becomes smaller. Extrinsic motivation is also similar to prior estimates, and at the higher end. The NR estimates closely resemble those in the baseline estimation, and the PR estimates resemble most closely those in the model with 39 moments. In estimates in the specification with $f = 15$ is somewhat different. As even low amounts of giving are assumed to suffice to generate positive image, the model tends to dismiss internal altruism as the main determinant of giving – it becomes significantly negative in all settings. The estimates of external motives, instead, become quite a bit larger, sometimes doubling in size. The main insight, however, is that the magnitudes remain comparable even under this somewhat stark assumption about external factors.
Finally, in column 7, we alter the $G$ and $H$ functions yet again to allow for sharing of more than half the endowment. The utility function now becomes

$$U_r(x) = x - (\alpha_r + \beta_r \mathbb{1}(\text{sort in}))x^2.$$  \hspace{1cm} (3)

This effectively redefines $\alpha$ and $\beta$ relative to a new maximum level of altruism, so we cannot compare the magnitudes of the estimates to the other specifications. Their relative levels (between treatments), however, are interpreted similarly. This change also requires a different breakdown of the data into moments, since a set of measure 0 is predicted to share exactly 5. We combine sharing exactly 5 and sharing more than 5, for a total of 27 moments instead of 33.

As a final step of the analysis, we contrast our estimates with those based a model that do not allow the extrinsic motivation to depend on the reciprocity environment ("Constant extrinsic motivation") and with estimates based on the traditional model ("No extrinsic motivation"). We use the model of column 1 in Table 6 and repeat the estimates in column 1 of Table 7 to ease comparison. Column 2 of Table 7 shows the estimates when extrinsic motivation is required to be constant across reciprocity environments, and column 3 shows the estimates when extrinsic motivation is required to be 0. As such, column 3 illustrates how well previous outcome-, type-, and intentions-based models of reciprocity fit the data.

In this estimation, to break ties, we assume that anyone indifferent between sorting out and sorting in but sharing nothing chooses to sort out. Within our general model assumption (indifferent non-sharers sort out with some fixed probability), this specification fits the data best and creates the most conservative comparison to the model with extrinsic factors.

Nonetheless, a model that does not allow for social pressure or other extrinsic factors does significantly worse at fitting the data. As shown in the column 3, the weighted SSE is much higher weighted SSE. The estimates are also rather implausible when applied to games with no social pressure. They imply that about half of all people will share in the DG (with sorting), and still 28% in a negative reciprocity environment (with sorting).

A model that allows for extrinsic motives but restricts them to be reciprocity-invariant, instead, yields estimates that are rather similar to those we obtain when we allow extrinsic motivation to vary. The estimates for the constant-beta model, shown in column 2, are rather close to those in column 1, both in the neutral (DG) and in the positive reciprocity settings. The estimates for the negative reciprocity case differ a bit more, with intrinsic motivation being estimated to be even more negative ($mu_\alpha = -7.561$ rather than $mu_\alpha = -5.723$) in order to counterbalance the constant and hence relatively high $\beta$. At the same time, standard errors are about halved, reflecting the gain in power from estimating fewer parameters.

Overall, the table reveals that allowing reciprocity to influence not only intrinsic but also extrinsic factors does not significantly improve the fit of the model, relative to assuming reciprocity-invariant extrinsic factors. This result parallels the reduced-form results that the impact of sorting on giving is approximately invariant across reciprocity environments.

We also vary the specification and assumptions of the model, as we did in Table 6. In all estimations, we replicate both patterns: requiring beta to be constant only slightly hurts the predictive power of the model, but requiring beta to be zero drastically reduces it. (We omit
these results for brevity.) We conclude that while reciprocity may slightly influence extrinsic motivations to share, it primarily acts through the channel of internal motives. Incorporating reciprocity-invariant social pressure into theories of reciprocity greatly improves their predictive power, but an additional interaction between extrinsic factors and reciprocity is relatively small on these scales.

At the same time, the estimation results also show that the amount of “additional” intrinsic motivation, in the sense of “reciprocity-induced” altruism or similar intrinsic motives under positive reciprocity is estimated approximately correctly under any of the models, including the naive model that does not allow for social pressure or other extrinsic motives. Specifically, if one asks how much “additional” intrinsic willingness to share is induced by a kind treatment of the recipient, the naive model implies that the average $\alpha_{NR}$ exceeds the average $\alpha_{DG}$ by 1.683. Under the models that allow for extrinsic motivation this difference between the $\mu_{\alpha}$'s amounts to 1.885 (variable beta) and 1.766 (constant beta), and is hence quite similar. This robustness reflects the fact that extrinsic motives are remarkably stable across both environments.

This conclusion is somewhat less true for the negative-reciprocity environment: The decrease in intrinsic willingness to share after an unkind treatment, relative to a neutral treatment, is slightly underestimated when neglecting extrinsic factors (the difference in $\mu_{\alpha}$'s amounts to −0.347) and is strongly overestimated when allowing only for a reciprocity-insensitive extrinsic motive (with a difference of 1.927 relative to the model reciprocity-dependent betas). In other words, failure to account for extrinsic motives and their context-dependence is particularly detrimental when estimating the motives for and extent of negative-reciprocal behavior.

For completeness, Table 8 shows the actual moments along with the predicted moments under the three models of Table 7. These detailed predictions mostly just break down the overall pattern of results described above, but we note two details that could lead to future refinements of reciprocity models. First, our distributional assumptions make it difficult for the model to match the fraction of small gifts, simply because the weight on the functions inducing sharing ($\alpha + \beta$) must lie between .1026 and .2105 for the dictator to give between 25 cents and $2.50, but only has to fall between .2105 and 4 to induce sharing between $2.75 and $4.75. Hence, it might be of value to explore alternative (more exotic) distribution or non-parametric estimations that allow for bunching to capture actual giving choices. In our contexts, standard distribution functions sufficed to generate robust results. Second, the data from the positive reciprocity conditions appears to fit the model less well than the other two reciprocity conditions, mainly due to the additional jump in small gifts up to $1. This could reflect a desire to return the $1 previously shared by the recipient, in line with Sugden’s (1984) view of reciprocity as an obligation to return favors at a minimal level. We omit such an ad hoc adjustment to the model for now, but a more detailed study of the function form of extrinsic motives for reciprocal sharing might require a more detailed exploration of this pattern.
6 Conclusion

In this paper, we question whether reciprocal behavior primarily reflect intrinsic motivations to share, such as altruism or fairness. Recent research has highlighted the importance of incorporating external factors, such as a concern for social image, into models of pro-social behavior. However, this recent research has, so far, had little impact on our notion and models of reciprocity.

We use a novel set of experimental conditions to show that reciprocal behavior responds to ‘avoidance opportunities’ in ways unaccounted for by existing models of reciprocity. The experimental conditions compare sharing behavior in a social environment where positive (or negative) reciprocity is in play to sharing under the same reciprocity condition but where the potential giver has the option to exit and avoid the sharing decision. We find that, regardless of the nature of the reciprocity environment, introducing an avoidance option causes generosity to drop significantly. The effect is comparable in size to the drop in a no-reciprocity (dictator game) environment. We also find that the rate of spitefully sorting in without sharing anything is decreasing in the inferred kindness of the recipient, i.e., out of proportion to the general rate of sharing zero. Traditional models of intrinsically motivated reciprocity—whether outcome-, type-, or intentions-based—are not able to account for these findings. That is, regardless of whether the choice context involves reciprocity or simple, one-sided sharing, extrinsic motivations are required to explain behavior.

We present a simple model that incorporates extrinsic determinants (e.g., social pressure) into a model of the spirit of previous reciprocity models, applicable to our experimental setting. By estimating this model, we demonstrate that even this simple extension of models of reciprocity is substantially better able to explain our results: Allowing for non-zero extrinsic motivations strongly improves the predictive power of the model. We also find, however, that reciprocity primarily changes behavior by changing people’s internal motivation to share, consistent with existing models.

As a next step in this research agenda, it would be interesting to conduct experimental treatment along the lines of the ones we use here to improve our understanding of reciprocity. For example, variations on standard reciprocity games, such as ultimatum games and the prisoner’s dilemma, similar to our sorting manipulation, could help identify preferences in a clean way and avoid mis-attributing behavior to easily confounded motivations. This approach should prove powerful in finalizing our understanding of reciprocal behavior. Similarly, field experiments are needed to explore whether reciprocity on real-world economic settings, such as contracts and labor markets, may in fact reflect extrinsic motives. For example, do above-market wages induce more effort due to intrinsic motives, such as increased altruism towards the employer, or might they also reflect social pressure to work hard when paid well?

As another important and bigger step, it will be important to pin down the type of preferences underlying our general term ‘extrinsic motivation.’ For example, it would be important to distinguish to what extent extrinsic determinants of sharing are social-image concerns, and to what extent givers succumb to social pressure. And, in the case of social image, is reciprocal giving welfare reducing, or is there an image reward for reciprocity?
Or do such motivations reflect self-image concerns, which can act similarly to social image motivations (as in, e.g., Bénabou and Tirole (2011) or Rabin (1995)), but might have different welfare implications? Moreover, if there is a confluence of factors, what is the net effect? DellaVigna et al. (2012) find that sharing decisions under social pressure are welfare-reducing, but it is possible that reciprocity of either variety counteracts this effect. It also remains to be discovered what factors trigger such extrinsic motivation.

Finally, we suggest that even the interpretation of the intrinsic determinant of reciprocal giving might need to be re-thought. Rather than caring about the intentions or type of the other person, the “intrinsic” element could reflect a sense of obligation in the sense of Cialdini (1993) and as proposed in the anthropology and sociology literature. This distinction that is particularly relevant for our economic understanding of the role of reciprocity in markets. When a firm gives gifts with the clear intention to maximize their profits, will consumers respond even if the firm’s intentions are clear? The literature addressing this question remains scarce and would benefit from further theoretical and empirical research.

References


Feld, Lars P. and Bruno S. Frey, “Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation,” Law & Policy, 2007, 29 (1), 102–120.


7 Tables

Table 1: $3 \times 2$ Experiment Design

<table>
<thead>
<tr>
<th>Standard Dictator Game</th>
<th>Double Dictator Game</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kind choice in initial mini-game</td>
</tr>
<tr>
<td></td>
<td>(No Reciprocity)</td>
</tr>
<tr>
<td>No Sorting</td>
<td>DG/NS</td>
</tr>
<tr>
<td>Sorting</td>
<td>DG/S</td>
</tr>
</tbody>
</table>
Table 2: Effect of Sorting on Sharing with Reciprocity

<table>
<thead>
<tr>
<th>Model:</th>
<th>OLS</th>
<th>Tobit</th>
<th>Probit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable:</td>
<td>Proportion Shared</td>
<td>Shared</td>
<td>Something</td>
</tr>
<tr>
<td>Sample</td>
<td>Baseline (No Sorting)</td>
<td>All Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.200***</td>
<td>0.193***</td>
<td>0.200***</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.024)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Negative Reciprocity</td>
<td>−0.130***</td>
<td>−0.105***</td>
<td>−0.130***</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.026)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Positive Reciprocity</td>
<td>0.039</td>
<td>0.438</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.031)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>Sorting</td>
<td>−0.063***</td>
<td>−0.079*</td>
<td>−0.135***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.043)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Sorting × Negative Reciprocity</td>
<td>0.040</td>
<td>0.042</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.119)</td>
<td>(0.157)</td>
</tr>
<tr>
<td>Sorting × Positive Reciprocity</td>
<td>0.011</td>
<td>0.041</td>
<td>−0.065</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.096)</td>
<td>(0.177)</td>
</tr>
<tr>
<td>Observations</td>
<td>99</td>
<td>283</td>
<td>283</td>
</tr>
<tr>
<td>(pseudo) $R^2$</td>
<td>0.113</td>
<td>0.145</td>
<td>0.147</td>
</tr>
</tbody>
</table>

Independent variables are condition dummies. The tobit model accounts for 147 observations being left-censored at zero. The probit model shows marginal effects. Robust standard errors are in parentheses (with bias-correction (HC3) in the linear case, see MacKinnon and White (1985)) and are calculated using jackknife estimation for the tobit model.

* - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$
Table 3: Rates of Sorting and Spiteful Non-sharing, Sorting conditions

<table>
<thead>
<tr>
<th>Model:</th>
<th>OLS</th>
<th>Probit</th>
<th>OLS</th>
<th>Probit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sorted Out</td>
<td>Spiteful Non-sharer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent Variable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.500***</td>
<td>0.109**</td>
<td>(0.075)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Positive Reciprocity (PR)</td>
<td>−0.183*</td>
<td>−0.186**</td>
<td>−0.077</td>
<td>−0.097*</td>
</tr>
<tr>
<td>Reciprocity (NR)</td>
<td>(0.096)</td>
<td>(0.094)</td>
<td>(0.052)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>PR-NR</td>
<td>0.087</td>
<td>0.087</td>
<td>0.091</td>
<td>0.072</td>
</tr>
<tr>
<td>(0.095)</td>
<td>(0.094)</td>
<td>(0.066)</td>
<td>(0.057)</td>
<td></td>
</tr>
<tr>
<td>PR-NR</td>
<td>−0.168***</td>
<td>−0.168***</td>
<td>(0.001)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Observations</td>
<td>184</td>
<td>184</td>
<td>184</td>
<td>184</td>
</tr>
<tr>
<td>(pseudo) $R^2$</td>
<td>0.055</td>
<td>0.041</td>
<td>0.050</td>
<td>0.076</td>
</tr>
</tbody>
</table>

Notes: “Sorted out” is a dummy variable for choosing the sorting out option in the sorting conditions; “spiteful non-sharer” is a dummy variable for a participant who chooses to sort in but shares nothing. Independent variables are condition group dummies. The probit model shows marginal effects. Robust standard errors are in parentheses (with bias-correction (HC3) in the linear case, see MacKinnon and White (1985).

* - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$
Table 4: **Zero, Low, and High Sharers: Effects of Reciprocity and Sorting**

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Zero sharers (1)</th>
<th>Low sharers (2)</th>
<th>Low sharers (3)</th>
<th>High sharers (4)</th>
<th>High sharers (5)</th>
<th>High sharers (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.384***</td>
<td>0.356***</td>
<td>0.293***</td>
<td>0.222***</td>
<td>0.323***</td>
<td>0.422***</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.073)</td>
<td>(0.046)</td>
<td>(0.063)</td>
<td>(0.048)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Positive Reciprocity</td>
<td>−0.240**</td>
<td>0.239**</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.120)</td>
<td>(0.126)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Reciprocity</td>
<td>0.323***</td>
<td>0.028</td>
<td>−0.351***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.117)</td>
<td>(0.106)</td>
<td>(0.091)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sorting</td>
<td>0.209***</td>
<td>0.253**</td>
<td>−0.048</td>
<td>−0.027</td>
<td>−0.160***</td>
<td>−0.227***</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.104)</td>
<td>(0.056)</td>
<td>(0.087)</td>
<td>(0.055)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>Sorting × Positive Reciprocity</td>
<td>−0.019</td>
<td>−0.054</td>
<td>0.073</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.148)</td>
<td>(0.150)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sorting × Negative Reciprocity</td>
<td>−0.145</td>
<td>−0.063</td>
<td>0.208*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.146)</td>
<td>(0.129)</td>
<td>(0.112)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>283</td>
<td>283</td>
<td>283</td>
<td>283</td>
<td>283</td>
<td>283</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.04</td>
<td>0.194</td>
<td>0.003</td>
<td>0.055</td>
<td>0.034</td>
<td>0.120</td>
</tr>
</tbody>
</table>

**Notes:** “Zero-sharers” is a dummy variable for either sorting out or sorting in but sharing nothing. “Low-sharers” is a dummy variable for sharing more than 0 but less than 25 percent of the endowment. “High-sharers” is a dummy variable for sharing at least 25 percent of the endowment. Independent variables are condition group dummies. Robust standard errors are in parentheses (with bias-correction (HC3), see MacKinnon and White (1985).

* - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$
Table 5: Model Predictions and Data

<table>
<thead>
<tr>
<th>1) Average giving</th>
<th>2) Sorting</th>
<th>3) Spiteful non-sharing</th>
<th>4) Impact of sorting on giving</th>
<th>5) Gift distributions sorting dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td>increasing in $r$</td>
<td>decreasing in $r$</td>
<td>decreasing in $r$</td>
<td>decreasing in $r$</td>
<td>no</td>
</tr>
</tbody>
</table>

No extrinsic motives: $\delta \alpha_i / \delta r > 0$, $\beta_i = 0$

Constant extrinsic motives: $\delta \alpha_i / \delta r > 0$, $\beta \sim F_\beta \forall r$

Variable extrinsic motives: $\delta \alpha_i / \delta r > 0$, $\delta \beta_i / \delta r > 0$

Data

<table>
<thead>
<tr>
<th>1)</th>
<th>2)</th>
<th>3)</th>
<th>4)</th>
<th>5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>yes</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>maybe</td>
<td>yes</td>
</tr>
</tbody>
</table>

Notes: Comparison of three restrictions on the distributions of $\alpha$ and $\beta$ in the model of Section 5.1.

1) Average giving is increasing in $r$: $\delta \bar{x} / \delta r < 0$.

2) Sorting is decreasing in $r$: $\delta P[U(0|\text{sort out}) > U(0|\text{sort in})]/\delta r < 0$.

3) Spiteful non-sharing is decreasing in $r$: $\delta P[U(0|\text{sort in}) > U(0|\text{sort out})]/\delta r < 0$.

4) Impact of sorting on giving is decreasing in $r$: $\delta(\bar{x}_{r,S} - \bar{x}_{r,NS})/\delta r < 0$.

5) Gift distributions are dependent on sorting: If $x_{r,S} \sim F_{x,r,S}$ and $x_{r,NS} \sim F_{x,r,NS}$, then $F_{x,r,S} \neq F_{x,r,NS}$.

“Yes” indicates that a prediction is required by a given model, “possible” indicates that the prediction is consistent with but not required by a model, and “no” means the model is not able to predict that aspect of the data. We assume here that $\alpha$ and $\beta$ have full support, so that even if with some distribution of types most people don’t change behavior when $r$ changes, as long as someone does, the model is still said to predict a change overall.
Table 6: Structural Estimation and Robustness Checks

<table>
<thead>
<tr>
<th>Specification</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictator</td>
<td>$\mu_\alpha$</td>
<td>-1.732 (0.486)</td>
<td>-2.214 (0.596)</td>
<td>-2.259 (0.562)</td>
<td>-2.819 (0.681)</td>
<td>-2.214 (0.556)</td>
<td>-3.937 (0.608)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\alpha$</td>
<td>3.569 (0.757)</td>
<td>4.171 (0.879)</td>
<td>3.389 (0.593)</td>
<td>3.907 (0.716)</td>
<td>3.147 (0.463)</td>
<td>1.422 (2.165)</td>
</tr>
<tr>
<td>Game</td>
<td>$\mu_\beta$</td>
<td>2.560 (0.489)</td>
<td>2.888 (0.560)</td>
<td>3.043 (0.573)</td>
<td>3.496 (0.661)</td>
<td>3.037 (0.555)</td>
<td>4.441 (0.603)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\beta$</td>
<td>3.159 (0.725)</td>
<td>3.646 (0.827)</td>
<td>3.057 (0.649)</td>
<td>3.589 (0.741)</td>
<td>3.492 (0.687)</td>
<td>4.244 (0.604)</td>
</tr>
<tr>
<td>Negative</td>
<td>$\mu_\alpha$</td>
<td>-5.723 (2.073)</td>
<td>-6.618 (2.499)</td>
<td>-8.866 (4.053)</td>
<td>-9.733 (4.670)</td>
<td>-5.787 (2.021)</td>
<td>-6.916 (2.233)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\alpha$</td>
<td>5.789 (2.124)</td>
<td>6.149 (2.234)</td>
<td>7.358 (3.352)</td>
<td>7.774 (3.693)</td>
<td>5.693 (1.921)</td>
<td>5.396 (2.705)</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>$\mu_\beta$</td>
<td>1.010 (0.657)</td>
<td>1.627 (0.846)</td>
<td>2.006 (0.965)</td>
<td>2.603 (1.106)</td>
<td>1.115 (0.747)</td>
<td>2.002 (1.141)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\beta$</td>
<td>1.627 (1.071)</td>
<td>2.670 (1.431)</td>
<td>2.507 (1.202)</td>
<td>3.267 (1.389)</td>
<td>1.748 (1.161)</td>
<td>2.944 (1.600)</td>
</tr>
<tr>
<td>Positive</td>
<td>$\mu_\alpha$</td>
<td>0.153 (0.140)</td>
<td>-0.063 (0.248)</td>
<td>0.110 (0.216)</td>
<td>-0.136 (0.358)</td>
<td>-0.773 (0.304)</td>
<td>-2.346 (0.773)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\alpha$</td>
<td>1.247 (1.005)</td>
<td>2.060 (0.992)</td>
<td>1.199 (1.048)</td>
<td>1.711 (1.120)</td>
<td>2.047 (0.191)</td>
<td>2.039 (2.102)</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>$\mu_\beta$</td>
<td>2.893 (0.450)</td>
<td>3.136 (0.503)</td>
<td>2.922 (0.404)</td>
<td>3.329 (0.461)</td>
<td>3.196 (0.492)</td>
<td>4.116 (0.997)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\beta$</td>
<td>3.485 (0.489)</td>
<td>4.003 (0.664)</td>
<td>2.891 (0.304)</td>
<td>3.469 (0.434)</td>
<td>3.455 (0.569)</td>
<td>3.436 (0.835)</td>
</tr>
<tr>
<td>Weighted SSE</td>
<td>288.671</td>
<td>279.286</td>
<td>284.820</td>
<td>276.454</td>
<td>287.512</td>
<td>296.274</td>
<td>53.142</td>
</tr>
</tbody>
</table>

Notes: GMM estimation results for baseline specification and six robustness checks. $\alpha$ refers to the weight on intrinsic motives, and $\beta$ the weight on extrinsic motives. Specification details (see section 5.1 for more details): (1) Baseline specification, 33 moments and normally distributed parameters. (2) 39 moments, normal distributions. (3) 33 moments, uniform distributions. (4) 39 moments, uniform distributions. (5) Alternative specification with $f = 5$ (see equation 2). (6) Alternative specification with $f = 15$ (see equation 2). (7) Alternative functions $G$ and $H$, with 27 moments (see equation 3).
Table 7: Structural Model Comparison

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_\alpha$</td>
<td>$\sigma_\alpha$</td>
<td>$\mu_\beta$</td>
</tr>
<tr>
<td>Dictator Game</td>
<td>-1.732 (0.486)</td>
<td>3.569 (0.757)</td>
<td>2.560 (0.489)</td>
</tr>
<tr>
<td></td>
<td>-1.643 (0.244)</td>
<td>3.395 (0.347)</td>
<td>2.435 (0.193)</td>
</tr>
<tr>
<td></td>
<td>3.159 (0.725)</td>
<td>6.031 (0.796)</td>
<td>3.372 (0.239)</td>
</tr>
<tr>
<td></td>
<td>$\sigma_\alpha$</td>
<td>$\mu_\beta$</td>
<td>$\sigma_\beta$</td>
</tr>
<tr>
<td></td>
<td>3.569 (2.073)</td>
<td>1.010 (0.657)</td>
<td>1.627 (1.071)</td>
</tr>
<tr>
<td></td>
<td>-7.561 (0.453)</td>
<td>6.922 (0.695)</td>
<td>3.372 (0.239)</td>
</tr>
<tr>
<td></td>
<td>-3.938 (1.064)</td>
<td>6.920 (1.423)</td>
<td></td>
</tr>
<tr>
<td>Negative Reciprocity</td>
<td>$\mu_\alpha$</td>
<td>$\sigma_\alpha$</td>
<td>$\mu_\beta$</td>
</tr>
<tr>
<td></td>
<td>0.153 (0.140)</td>
<td>1.247 (1.005)</td>
<td>2.893 (0.450)</td>
</tr>
<tr>
<td></td>
<td>0.123 (0.211)</td>
<td>1.961 (0.276)</td>
<td>2.435 (0.193)</td>
</tr>
<tr>
<td></td>
<td>1.389 (0.352)</td>
<td>5.220 (0.557)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma_\alpha$</td>
<td>$\mu_\beta$</td>
<td>$\sigma_\beta$</td>
</tr>
<tr>
<td></td>
<td>1.247 (1.005)</td>
<td>2.893 (0.450)</td>
<td>3.485 (0.489)</td>
</tr>
<tr>
<td></td>
<td>1.961 (0.276)</td>
<td>2.435 (0.193)</td>
<td>3.372 (0.239)</td>
</tr>
<tr>
<td>Positive Reciprocity</td>
<td>$\sigma_\beta$</td>
<td>$\mu_\beta$</td>
<td>$\sigma_\beta$</td>
</tr>
<tr>
<td></td>
<td>2.893 (0.450)</td>
<td>2.435 (0.193)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.485 (0.489)</td>
<td>3.372 (0.239)</td>
<td></td>
</tr>
<tr>
<td>Weighted SSE</td>
<td>288.671</td>
<td>291.839</td>
<td>380.243</td>
</tr>
</tbody>
</table>

**Notes:** GMM estimation results for baseline model specification (33 moments, $\alpha$ and $\beta$ normally distributed) and comparison models requiring that social pressure be invariant to reciprocity, or zero social pressure. $\alpha$ refers to the weight on internal altruism, and $\beta$ the weight on external social image.
## Table 8: Moments and Model Predictions

<table>
<thead>
<tr>
<th>Dictator Game</th>
<th>Actual</th>
<th>Baseline Model</th>
<th>Constant Social Pressure</th>
<th>No Social Pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sort in, give 0</td>
<td>0.1087</td>
<td>0.1725</td>
<td>0.1973</td>
<td>0.0000</td>
</tr>
<tr>
<td>Sort in, give 0.25 – 2.50</td>
<td>0.2174</td>
<td>0.0015</td>
<td>0.0017</td>
<td>0.0071</td>
</tr>
<tr>
<td>Sort in, give 2.75 – 4.75</td>
<td>0.0435</td>
<td>0.0911</td>
<td>0.0924</td>
<td>0.2284</td>
</tr>
<tr>
<td>Sort in, give 5</td>
<td>0.1087</td>
<td>0.1769</td>
<td>0.1686</td>
<td>0.2382</td>
</tr>
<tr>
<td>Sort out, give &gt; 5</td>
<td>0.0217</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative Reciprocity</th>
<th>Sort Out</th>
<th>0.5000</th>
<th>0.5580</th>
<th>0.5399</th>
<th>0.5262</th>
</tr>
</thead>
<tbody>
<tr>
<td>Give 0</td>
<td>0.3556</td>
<td>0.4395</td>
<td>0.4428</td>
<td>0.5262</td>
<td></td>
</tr>
<tr>
<td>Give 0.25 – 2.50</td>
<td>0.2667</td>
<td>0.0089</td>
<td>0.0089</td>
<td>0.0071</td>
<td></td>
</tr>
<tr>
<td>Give 2.50 – 4.75</td>
<td>0.2000</td>
<td>0.2986</td>
<td>0.2971</td>
<td>0.2284</td>
<td></td>
</tr>
<tr>
<td>Give 5</td>
<td>0.1778</td>
<td>0.2529</td>
<td>0.2512</td>
<td>0.2382</td>
<td></td>
</tr>
<tr>
<td>Give &gt; 5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive Reciprocity</th>
<th>Sort Out</th>
<th>0.0317</th>
<th>0.1689</th>
<th>0.1824</th>
<th>0.0000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Give 0</td>
<td>0.3968</td>
<td>0.0034</td>
<td>0.0033</td>
<td>0.0080</td>
<td></td>
</tr>
<tr>
<td>Give 0.25 – 2.50</td>
<td>0.1111</td>
<td>0.1723</td>
<td>0.1702</td>
<td>0.2808</td>
<td></td>
</tr>
<tr>
<td>Give 2.50 – 4.75</td>
<td>0.1270</td>
<td>0.2463</td>
<td>0.2505</td>
<td>0.3085</td>
<td></td>
</tr>
<tr>
<td>Give &gt; 5</td>
<td>0.0159</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

| Weighted SSE | 288.7 | 291.8 | 380.2 |

**Notes:** Actual and predicted moments from GMM estimations of baseline model, model requiring that social pressure be invariant to reciprocity, and model requiring no social pressure.
8 Figures
Figure 1: Distributions of Amounts Shared

(a) DG condition: No Reciprocity

(b) PR condition: Positive Reciprocity

(c) NR condition: Negative Reciprocity
Figure 2: Average Amounts Shared with and without Reciprocity

Figure 3: Sorting In to Share Something or Nothing
Figure 4: Parameterization of Reciprocity Model