Review in Advance first posted online on May 9, 2014. (Changes may still occur before final publication online and in print.)

Rethinking Reciprocity

Ulrike Malmendier,[1] Vera L. te Velde,[1] and Roberto A. Weber[2]

[1]Department of Economics, University of California, Berkeley, California 94720; email: ulrike@berkeley.edu

[2]Department of Economics, University of Zurich, Zurich CH-8006, Switzerland

Keywords

social preferences, social pressure, social image, external and internal motivations, structural estimation, double-dictator game

Abstract

Reciprocal behavioral has been found to play a significant role in many economic domains, including labor supply, tax compliance, voting behavior, and fund-raising. What explains individuals' tendency to respond to the kindness of others? Existing theories posit internal preferences for the welfare of others, inequality aversion, or utility from repaying others' kindness. However, recent evidence on the determinants of (unilateral) sharing decisions suggests that external factors such as social pressure are equally important. So far, this second wave of social preference theories has had little spillover to two-sided reciprocity environments, in which one individual responds to the actions of another. We present a novel laboratory reciprocity experiment (the double-dictator game with sorting) and show that failure to account for external motives leads to a significant overestimation of internal motives such as fairness and altruism. The experimental data illustrate the importance of combining reduced-form and structural analyses to disentangle internal and external determinants of prosocial behavior.

Changes may still occur before final publication online and in print

1. INTRODUCTION

Reciprocal behavior is of increasing interest in many areas of economics. Labor economists argue that firms pay above-market wages to induce reciprocal behavior among their workers (Akerlof 1982, Bewley 2009, Fehr et al. 2009), which in turn might explain involuntary unemployment (Akerlof & Yellen 1988, 1990) and inflation (Okun 1981). Conversely, workers who feel mistreated by their employer exert negative reciprocity and lower effort or produce faulty products (Krueger & Mas 2004, Kube et al. 2013). 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 & Frey 2007). Political economists provide evidence that politicians target pre-election transfers toward reciprocal individuals to increase their vote shares (Finan & Schechter 2012). And health economists show that pharmaceutical representatives exploit the reciprocity of doctors with small gifts (Brennan et al. 2006). Much of this literature in economics builds on earlier work in psychology, including the seminal work by Cialdini (1993) on the use of small gifts in marketing and politics with the aim to trigger reciprocity.

Given that reciprocity appears to affect a broad class of market interactions, market features such as dynamic pricing strategies (e.g., low introductory prices), consumer loyalty, and employee turnover are in turn likely to be affected by social preferences. Profit-maximizing firms need to anticipate and respond to the nonstandard behavior of their consumers.

Despite the mounting empirical evidence on the importance of reciprocal behavior, economists still struggle to converge on the correct model of the underlying motives. Existing theories rely on the response to good intentions or the kindness of others as the trigger for reciprocal behavior (e.g., Rabin 1993, Levine 1998). Under these theories, people reciprocate because another person's kind act or benevolent nature increases the intrinsic utility of acting kindly toward this person. Thus, such preferences are internal in that they arise from an individual's preference to act in a way that rewards good behavior by others.

We propose that the analysis of reciprocity needs to be rethought. Our argument is motivated by a recent literature that revisits the motives for sharing in simpler, one-sided settings (without reciprocity). This recent literature reveals that seemingly altruistic people are often reluctant to share and avoid the opportunity to share if they can (Dana et al. 2006, Broberg et al. 2007, DellaVigna et al. 2012, Lazear et al. 2012). Such evidence has led to a revision of the conventional theoretical motives thought to underlie generous or charitable behavior: Not only internal preferences for equality or social welfare, but also external factors such as social pressure, social image, and norms appear to be important determinants of seemingly altruistic behavior (Harbaugh 1998, Akerlof & Kranton 2000, Bénabou & Tirole 2006, Battigalli & Dufwenberg 2007).[1]

These advances in our understanding of one-sided giving decisions have thus far had little spillover to our understanding of two-sided prosocial behavior, in which the decision to give is a function of another person's earlier behavior. Omitting external factors in reciprocity settings could bias our understanding of social preferences. Consider the following example. Suppose that, in a simple dictator game (DG), 65% of dictators share. Now allow recipients to send a small gift to the dictators, prior to the DG, and suppose that the percentage of dictators who share increases to 80%. A simple model of internally motivated reciprocity would attribute the decision of all 80% of sharers to welfare-enhancing motivations. This interpretation might be wrong on two counts. First, some of the 65% who share in both the simple one-sided DG and the two-sided reciprocity

---

[1]We use the terms internal and external to refer to the primary source of the motivation as coming from inside the individual (preferences) or from outside (social considerations).

setting might be motivated by external factors, such as social pressure or social image. Second, some of the additional 15% who share only in the two-sided game may be responding to external motives (such as an increase in social pressure) rather than changes in internal motives (such as an increase in altruism induced by the recipient's prior act of kindness).

Our review discusses the existing literature on reciprocity, including the leading theoretical models, and points to the common failure among these models to account for the possibility of external motives. We then describe a novel experiment mirroring the above example that explores the importance of internal versus external determinants of reciprocal behavior. We employ a double-dictator game (DDG), which allows us to compare giving in a positive and a negative reciprocity setting with giving in a neutral, one-sided setting (the standard DG). We modify these situations by giving dictators the option to avoid the sharing decision. The 3×2 design (neutral/positive/negative $\times$ no-sorting/sorting) allows us to disentangle internal and external factors determining giving in one-sided versus two-sided (reciprocity) giving environments.[2]

The findings cast a new light on the motives underlying reciprocity. The experiment reveals that, even after inducing positive reciprocity, external motives are still at work, approximately to the same extent as in the simple (neutral) DG setting. Although a prior gift increases giving in the DDG, relative to the baseline DG, the increase is not robust to giving subjects the option to avoid reciprocal behavior. When subjects can opt out in the positive reciprocity setting, they do so, approximately to the same extent to which subjects in the simple DG opt out. Hence, external factors are important determinants of sharing in reciprocity environments as well. The two motives appear to be approximately additive. Failure to account for external motives leads to a significant overestimation of internal motives such as fairness and altruism in positive reciprocity settings. We also find that negative reciprocity virtually eliminates giving in the setting with sorting and even induces spite: Some people sort in and share zero.

We confirm these conclusions in a simple structural model that distinguishes between internal and external motives. Even after the inducement of positive reciprocity, we estimate a significant influence of external factors, similar in size to the strength of external factors in one-sided (standard) DGs. We compare our estimates to those based on a naïve structural model that does not account for external motives. Consistent with the reduced-form results, failure to account for external motives leads to a sizable overestimation of internal motives, under both positive and negative reciprocity. At the same time, the estimation results also confirm that the amount of additional internally motivated sharing is estimated approximately correctly in the naïve model. In other words, it is correct to attribute the additional giving in a positive reciprocity environment, compared to a neutral setting, to internal preferences, and even a naïve model delivers the correct estimate. This robustness reflects that external motives are stable across both environments.[3]

In summary, the data and analyses support the broader point of this review: External motives are a significant determinant of prosocial behavior in reciprocity environments, similar in strength to their role in one-sided environments without reciprocity. This insight is particularly relevant for the analysis of social preferences in markets. It suggests that market participants aiming to trigger positive reciprocity need to account for observability and avoidance options. For example,

---

[2]We note that external factors are not necessarily welfare reducing; i.e., avoidance is sufficient but not necessary to identify externally motivated giving. For example, social image gains might increase a giver's net utility. We focus on welfare-reducing external determinants, such as social pressure, for ease of identification.

[3]This conclusion is somewhat less true for the negative reciprocity environment. Our full-model structural estimates reveal instead that individuals feel less external pressure to share after having been treated unkindly. In other words, failure to account for context-dependent external motives is particularly detrimental when estimating the motives for and extent of negative reciprocal behavior.

a charity might benefit from giving a small gift to the potential donor, as shown by Falk (2007), but even after receiving the gift, some donors will give because they feel pressure or obligations to give, not because they enjoy giving. Hence, the gift will be most effective if the recipient cannot avoid the subsequent request for a donation.

A second main insight relates to the value of combining reduced-form and structural analyses in explaining prosocial behavior when the outcome (sharing) could reflect rather different psychological motives. The structural estimation allows us to decompose the share of giving that results from internal versus external determinants, both in one-sided and in two-sided giving contexts.

We organize the remainder of the article 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 external 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 describe the results of a novel experiment that allow us to better understand the motives underlying reciprocal behavior (Section 3). We discuss a simple model integrating internal and external 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 for economic decisions has been a source of debate in the profession for decades. How do we explain sharing behavior in standard economic settings when it is inconsistent with a classical model of self-interested preferences? In this article, we focus on what one might call two-sided prosocial behavior, in which sharing is a response to the other person's prosocial behavior. However, our motivation for rethinking the reciprocity literature stems from recent advances in the literature on one-sided sharing decisions, in which one party shares with another party without consideration of the other party's (prior) sharing behavior. We first provide a perspective on these recent advances.

2.1. One-Sided Giving

A large volume of laboratory and field experiments documents voluntary sharing behavior and studies the individual characteristics and contextual factors that influence sharing. These studies show that generosity varies based on personal characteristics such as the gender of the decision maker, the framing of the decision, the source of the surplus, and the social context in which an altruistic choice is embedded (Hoffman et al. 1996, Andreoni & Vesterlund 2001, Andreoni & Miller 2002, Cherry et al. 2002, Fong & Luttmer 2009, Henrich et al. 2010, Brock et al. 2013, DellaVigna et al. 2013). Although 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 & Miller 2002) or in the form of maximin preferences (Charness & Rabin 2002), as well as different specifications of inequality aversion (Fehr & Schmidt 1999, Bolton & Ockenfels 2000). These models assume 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 making an explicit decision and to thereby keep their endowment (Dana et al. 2006, Broberg et al. 2007). The novel

36.4   *Malmendier • te Velde • Weber*

design feature in these studies is to give the decision maker the option of avoiding the decision to give. For example, in the laboratory study of Lazear et al. (2012), subjects are allowed to opt out of playing the DG. Or, in the field study of DellaVigna et al. (2012), the exact time of a door-to-door fund-raiser is preannounced so that people can choose not to open the door of their homes. In both cases, most people avoid the socially unpleasant 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 prosocially. For example, Dana et al. (2007) find that giving decreases when the recipient does not find out with certainty how a choice was made or when the dictator does not know how his choice will affect the recipient.[4] In fact, dictators prefer not to find out about the effect of their choice. Andreoni & Bernheim (2009) manipulate the probability that a computer secretly overrides the dictator's choice with a known default, showing that dictators will take the opportunity to hide their selfish actions when possible. Grossman (2012) uses a similar probabilistic DG to demonstrate the importance of social signaling relative to self-signaling. Moreover, several of the experiments mentioned above demonstrate that people prefer to avoid letting the recipient know about the game, even at a cost to the player hiding the information (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, 1996) show that DG giving drops when using a double-blind framework. Additionally, Bohnet & Frey (1999a,b) show that giving increases when the dictator and recipient face each other. Relatedly, Franzen & Pointner (2012) show that DG sharing is reduced when choices are concealed from the experimenter using a randomized response technique.

Outside of the laboratory, anonymity and observability appear to play a similar role. For example, Mas & Moretti (2009) show that store employees work harder when observed by other employees 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 a charity. In other words, generosity decreases significantly when individuals can avoid having to say no directly 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, much voluntary sharing appears to stem from the difficulties individuals have in saying no to a request and from 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 & Tirole 2006, Battigalli & Dufwenberg 2007, Andreoni & Bernheim 2009).

In summary, the above stream of social preference research can be organized into two waves: (*a*) empirical evidence that people act in a manner consistent with a preference for altruism or

---

[4]A more extreme way to remove transparency is to make the recipient entirely unaware of a game. Dana et al. (2006) show that dictators are less generous in this case. Koch & Normann (2008) and Johannesson (2000) instead find similar generosity in a standard DG and a variant with uninformed recipients. It remains to be shown whether differences in experimental designs (e.g., double-blind anonymity, the precise types of recipients) account for the difference in findings.

fairness, and theoretical models of such internal motivations, and (*b*) evidence that such behavior is susceptible to avoidance options and variations in observability, along with a new class of social preference theories that account for such external motivations.

2.2. Two-Sided Giving

So far, the new insights of the second wave of literature have largely failed to spill over to a broader set of social preferences beyond 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 & Rabin 2002, Falk & Fischbacher 2006). A preference for acting reciprocally has been advanced as underlying many puzzling phenomena, both in the field, as discussed in Section 1, 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 & Gächter 2000b for a review). The pattern of reciprocal behavior survives in one-shot, anonymous scenarios (Hoffman et al. 1998) and persists over time and through learning opportunities (Keser & van Winden 2000). People even reciprocate on behalf of others (Carpenter & Matthews 2004, Carpenter et al. 2004, Fehr & 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 the response is strong enough to yield a nonnegative 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 & Gächter 2000a, Ostrom et al. 1992). Croson (2007) uses variants of the public good game to show that reciprocity is a stronger motivation than is either altruism or the moral commitment to act in the manner one would prefer everyone to act.

Evidence from ultimatum game experiments conducted with populations from multiple countries shows that reciprocity is a cross-cultural phenomenon that persists even when the stakes are as high as several months' wages (Camerer & Thaler 1995, Roth & Erev 1995, Henrich et al. 2001). A critical element driving reciprocity in such games seems to be the intention and agency behind the initial act (Blount 1995, Brandts & Solà 2001). Thus, in the ultimatum game, reciprocity also manifests itself 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 noncontractible. 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 do the laboratory experiments but confirm its existence and help to identify the conditions under which gift exchange occurs. For example, Gneezy & List (2006) find that workers reciprocate surprise gifts with higher effort, at least initially (see also List 2006). Kube et al. (2012) show that the type of gift is important: Small nonmonetary gifts and monetary gifts presented thoughtfully are better at inducing positive reciprocity. Falk (2007) shows that donation requests are returned more frequently when they include a token gift (a postcard).

With regard to negative reciprocity, Greenberg (1990) documents a rise in employee theft after wage cuts, and Krueger & 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. (2013) argue that negative reciprocity—induced by a surprise wage cut—is stronger and more robust than is 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, and intentions-based models.[5] We discuss these models here with an eye toward our main argument—the relevance of external determinants of reciprocal behavior—and the particular experimental analysis of a one-shot, nonstrategic, anonymous game, to be discussed in the next two sections.[6]

Outcome-based models 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, in outcome-based models, reciprocity is the result of agents trying to rebalance their allocations in response to a kind or unkind transfer from others.

In the category of outcome-based models, one model that can produce reciprocity is pure altruism. Altruists incorporate others' material outcomes into their utility just like any other good to which they wish to allocate wealth. So when another's material outcome is reduced by a costly kind act, an altruist may in turn desire to help restore the other's material wealth. Simple forms of altruism have 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 & Miller (2002) formulate a model of utilitarianism that has been widely used in the social preference literature. Impure altruism, also known as warm glow (Andreoni 1989, 1990), works similarly in that a giver obtains direct utility from acting to improve another person's material outcome. In this case, givers do not care directly about the outcome (e.g., total donations, including from other sources) but care about their personal act of giving.

Another class of outcome-based models that can generate reciprocal behavior assumes distributional preferences. One example involves maximin preferences, as in Charness & Rabin (2002), for which a giver would like the minimum payoff anybody receives to be as large as possible.[7] Another example is inequality aversion, as modeled by Fehr & Schmidt (1999) and Bolton & Ockenfels (2000). All 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, as suggested by Ledyard (1997) and used by Cappelen et al. (2007).

In the category of type-based models, also known as interdependent preferences, the seminal model of Levine (1998) allows the weight someone places on another person's material outcome to depend both on a personal altruism parameter and on the other person's altruism parameter. That is, nice people are nicer to others in general, and everyone is nicer to nice people. On the opposite

---

[5]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 reciprocal using bounded rationality (Gale et al. 1995, Roth & Erev 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, Bowles & Gintis 2011).

[6]Sobel (2005) and Cooper & Kagel (2014) provide excellent additional discussions of reciprocity.

[7]Charness & Rabin (2002) also allow an agent to place less weight on others' payoffs after they have misbehaved, thus incorporating an element of the type-based models discussed next.

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 people according to how they have previously treated others).

In the category of intentions-based models, the foundational example is Rabin's (1993) psychological equilibrium model, in which beliefs enter directly into utility (see also Geanakoplos et al. 1989, Battigalli & Dufwenberg 2009). 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 other's actions (and second- and third-order beliefs) must be correct, and players must optimally respond to those actions and to the kindness 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. This model diverges from models such as Fehr & Schmidt (1999) and Levine (1998) in that people directly care about treating others the way they expect 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 applies only to two-person normal form games with pure strategies. Others have developed more general versions of similar models (Dufwenberg & Kirchsteiger 2004, Falk & Fischbacher 2006). Such intentions-based models have been corroborated experimentally by, for example, Dhaene & Bouckaert (2010), who show that Dufwenberg & Kirchsteiger's (2004) model is consistent with a large majority of people's behaviors 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.

2.3. The Missing Piece

One way to frame the limitation of these 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 reviewed above, the same limitation applies 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 external factors, such as social pressure and social image.

Research on reciprocity has largely not yet incorporated these more recent approaches.[8] The neglect of external factors could bias our understanding of reciprocity, as demonstrated with the following experimental test.

3. EXTERNAL MOTIVES IN RECIPROCITY: AN EXPERIMENTAL TEST

Whether sharing in reciprocity environments is fully explained by internal 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. Here, we apply Lazear et al.'s (2012) experimental approach, which allows agents to avoid making a decision and keep all of their wealth, to a reciprocity setting. If reciprocal giving reflects only internal motives, avoidance options should not affect giving behavior.

3.1. Experimental Design

In a standard DG, the decision maker has no prior interaction with the recipient. Hence, reciprocity is irrelevant, and the dictator's choice involves only trade-offs between personal material payoffs, internal motivations to share (e.g., altruism), and external motivations to share (e.g., social pressure). In order to test whether external determinants play a role in reciprocal behavior, we use variants of a DDG. Here, the dictator and recipient play a mini-DG over $2 prior to the main DG (over $10), with the dictator and recipient roles switched.[9] The purpose of the initial reversed mini-DG is to induce reciprocity as a motive for sharing in the second-stage main DG. The mini-dictator is given a choice between sharing $0 or $1. This binary choice allows clean assignment to reciprocity treatments: Positive reciprocity is induced if the mini-dictator shares $1, and negative reciprocity is induced if the mini-dictator decides not to share. (The standard DG is the neutral no-reciprocity benchmark for comparison.[10]) The parties learn about the second-stage DG only after the mini-game has been played and the results have been revealed to the other party.[11]

We cross these three reciprocity conditions (neutral DG and positive and negative reciprocity DDGs) with a sorting option that allows the (main) dictator to avoid the sharing decision and keep the endowment. In the no-sorting condition, the dictator is forced to choose an allocation, of which the recipient is then informed; in the sorting condition, 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.

Full implementation details for the experiment are provided in **Supplemental Appendix A** (follow the Supplemental Material link from the Annual Reviews home page at **http://www. annualreviews.org**). In brief, we employed a between-subjects design for all 3×2 treatment conditions. All experiments were conducted at the University of California, Berkeley, with a total of 192 pairs of subjects in the DDGs and 91 pairs in the DGs. Dictator and recipient roles were

---

[8]One recent exception is given by van der Weele et al. (2014). They provide subjects with a means to plausibly deny responsibility for failing to act reciprocally. This manipulation has little effect on reciprocal behavior.

[9]The smaller stakes in the first-stage mini-DG allow us to distinguish reciprocity from distributional preferences such as inequity aversion.

[10]With this comparison, we avoid confounds with other social preferences, as cautioned by Cox (2004).

[11]Initial voluntary sharing can thus be interpreted as an act of kindness, rather than as an attempt to induce reciprocal behavior, which avoids concerns about interpreting the dictators' reactions as reciprocity.

randomly assigned, and the two types were moved to separate rooms at the beginning of the experiment, before receiving instructions.

These six conditions—three reciprocity conditions, each with and without a sorting opportunity—allow us to evaluate the role of internal and external determinants of sharing. As described above, the leading models of reciprocity attribute giving to internal motives, and the sorting option should be irrelevant after positive reciprocity has been induced. Thus, although we know that about half of the dictators who share in a simple DG prefer to sort out if possible, the existing models of reciprocity predict no such sorting behavior after inducing reciprocity. Similarly, negative reciprocity induced by not sharing in the prior mini-game should induce spite and a willingness to punish the other party, which make people happy to share nothing regardless of whether they can opt out.

Alternatively, external motives may be at play in reciprocity environments. For example, positive reciprocity may increase internal motivation but not eliminate the external motivation we have observed in no-reciprocity settings. In this case, we would expect to see some sorting out in the DDG after the mini-dictator has shared. Another possibility is that external motives affect reciprocal behavior directly. For example, positive reciprocity could increase the image cost of being ungenerous. In this case, the prediction regarding opt-out frequency is ambiguous. With both internal and external determinants becoming stronger, it could increase or decrease as sorting in is now more costly (owing to increased social pressure), and sorting out is more costly (owing to increased internal motivation). However, we do have a clear prediction for average sharing conditional on not opting out: Those deciding to play the (main) DG will share more. We further explore these possibilities in the context of a formal model in Section 5.

3.2. Experimental Results

Figure 1 shows the distribution of the amounts shared in each condition, with the frequencies of opting out separately at the left. **Table 1** presents the corresponding reduced-form regression analyses, which also indicate the mean amounts shared in each condition. More details and supporting analyses are in **Supplemental Appendix B**.

We first consider the effect of reciprocity on sharing without any sorting options. In the standard DG, most subjects share a positive amount (64%). The mean amount shared is $2.00 out of $10.00. Average sharing rises to $2.39 in the positive reciprocity condition, and it falls to $0.70 in the negative reciprocity condition of the DDG. Column 1 of **Table 1** shows that the decrease in negative reciprocity (−$1.30) is significant but that the increase in positive reciprocity is not (+$0.39),[12] consistent with previous experimental evidence (e.g., weak positive reciprocity but strong "concern withdrawal" in Charness & Rabin 2002). The above averages are, however, based on the second-stage endowment of $10. If we add the $2 from the mini-DG to the total endowment (see Cox 2004), the positive reciprocity effect is marginally significant, and the negative reciprocity effect is insignificant.[13] The subsequent analysis on the effects of sorting is unchanged when $12 is used as the relevant endowment.

---

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

[13]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 DG. Considering the percentage obtained by the recipient, 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 the recipients kept the initial $2.
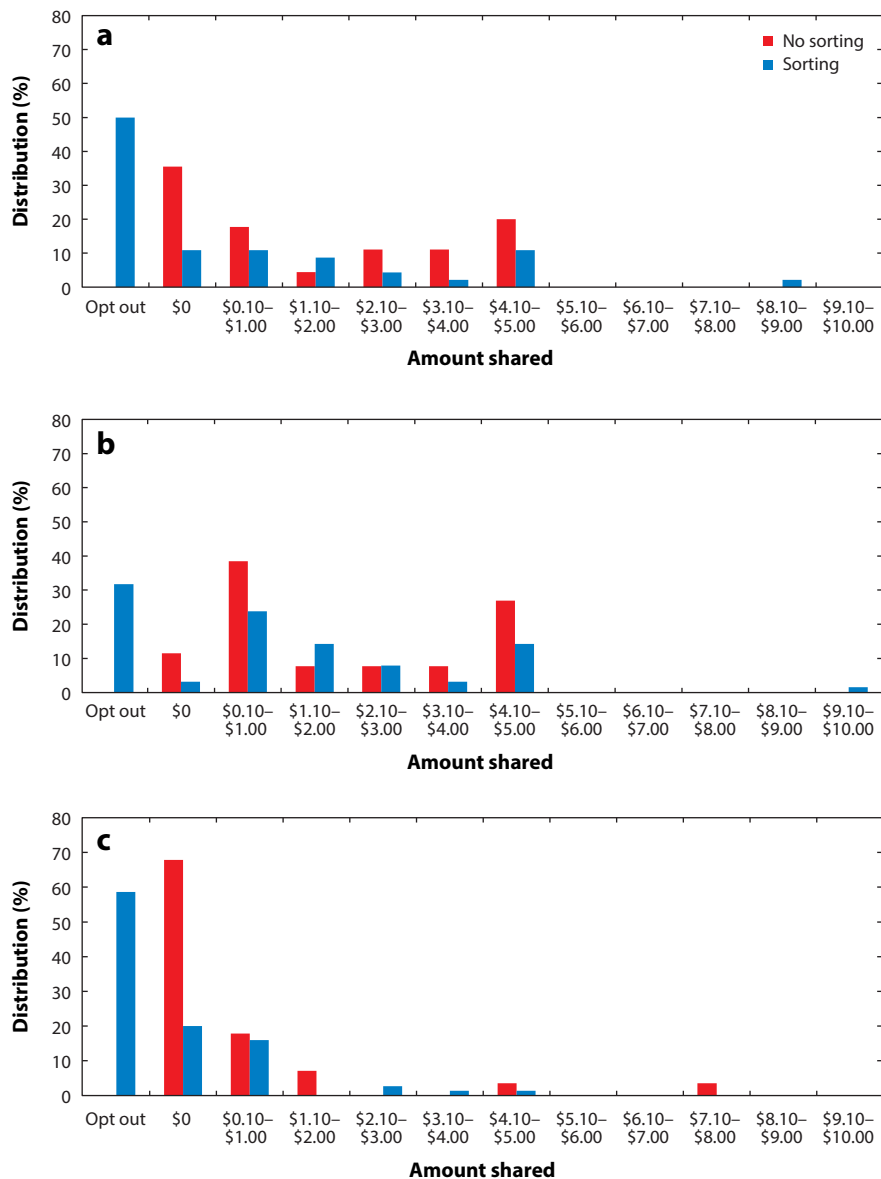
Figure 1

Distributions of amounts shared in each condition in the dictator game (DG) and double-dictator game (DDG): (*a*) the DG condition with no reciprocity, (*b*) the positive reciprocity (PR) condition of the DDG, and (*c*) the negative reciprocity (NR) condition of the DDG.

3.2.1. The effect of sorting on sharing. The introduction of sorting strongly decreases the average amount shared (to $1.21) and the frequency of sharing (to 39%) in the DG. In the positive reciprocity treatment, sorting also causes a large drop in average amounts shared, to $1.71. The amount shared decreases to $0.31 in the negative reciprocity treatment. **Table 1** confirms the statistical significance of the sorting-induced decrease in the amount shared, both in a linear regression (columns 1–3) and in a Tobit specification (columns 4 and 5).

Table 1 Effect of sorting on sharing with reciprocity

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS			Tobit		Probit
Dependent variable	Proportion shared			Proportion shared		Shared something
Constant	0.200*** (0.030)	0.193*** (0.024)	0.200*** (0.030)	0.119*** (0.041)	0.131*** (0.048)	
Negative reciprocity	−0.130*** (0.045)	−0.105*** (0.026)	−0.130*** (0.045)	−0.234*** (0.056)	−0.257*** (0.090)	−0.319*** (0.111)
Positive reciprocity	0.039 (0.049)	0.438 (0.031)	0.039 (0.049)	0.112** (0.049)	0.089 (0.066)	0.319** (0.13)
Sorting		−0.063*** (0.024)	−0.079* (0.043)	−0.135*** (0.042)	−0.161** (0.074)	−0.253** (0.101)
Sorting × negative reciprocity			0.040 (0.055)		0.042 (0.119)	0.125 (0.157)
Sorting × positive reciprocity			0.011 (0.063)		0.041 (0.096)	−0.065 (0.177)
Observations	99	283	283	283	283	283
(Pseudo-)R^2	0.113	0.145	0.147	0.209	0.209	0.150

Independent variables are condition dummies. Column 1 uses only the dictator game data, and all other columns use both dictator game and double-dictator game data. 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 & White 1985] and are calculated using jackknife estimation for the Tobit model. $p < 0.1$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

The dark bars in **Figure 1** illustrate the sharp shift of the distributions to the left when sorting becomes possible, regardless of the reciprocity conditions. In fact, when categorizing choices as sharing nothing, a small amount, or a generous amount, we find a significant drop in generous sharing and a significant increase in sharing nothing when sorting is introduced.[14] This is true in all three reciprocity treatments, and the magnitude of the effect is not significantly different between reciprocity treatments.

Hence, both the simple comparison of means and the distributional evidence suggest that sorting has a large impact on sharing, even under reciprocity. 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 suggests that the dominant approach to modeling sharing under reciprocity, which relies on internal factors, is incomplete. External factors affect individuals' giving in reciprocity environments as well.

But how strong are external motives after the inducement of positive or negative reciprocity? Is their role similar to the one found in neutral (one-sided) settings? The simple comparison of means suggests that the effect is smaller under positive reciprocity and larger under negative reciprocity

---

[14]Readers are referred to **Supplemental Appendix B** for details. Midlevel sharing does not change significantly. Without within-subject data, we cannot infer whether there is an overall shift to the left or whether high sharers convert to zero sharers when sorting is an option. The structural analysis of Section 5 sheds light on this issue.

than it is in the neutral DG setting, although the differences do not seem large. Sharing drops by 40% in the standard DG when sorting is introduced, by 29% in the positive reciprocity condition, and by 56% in the negative reciprocity condition. Columns 3 and 5 of **Table 1** reveal that the sorting-induced decrease in sharing does not differ significantly across the DG and the (positive or negative reciprocity) DDG conditions. The same picture emerges if we consider the frequency of sharing. The probit regression in the final column of **Table 1** shows that 25% of sharers opt out, but the interactions of sorting with either reciprocity condition, positive or negative reciprocity, are statistically insignificant.

In other words, the impact of sorting on average amounts shared is large and significant, and is approximately invariant to the reciprocity setting. We investigate these reduced-form findings in more detail in our structural estimation in Section 5.2, in which we provide quantification of the role of internal and external factors.

3.2.2. The effect of reciprocity on sorting.
The above analysis of amounts shared suggests that external factors affect reciprocal behavior. Otherwise, giving would have been unaffected by the opportunity to avoid the giving request. We now test directly to what extent the reduction in giving comes via the channel of opting out.

The leftmost bars in each panel of **Figure 1** reveal that the sorting option is used by a significant fraction of subjects in all treatments: 50% in the DG, 32% in the positive reciprocity condition of the DDG, and 59% in the negative reciprocity condition of the DDG. The differences in frequencies are either insignificant (negative reciprocity versus DG) or only marginally significant (positive reciprocity versus DG), as described in **Supplemental Appendix B**. This evidence speaks to the presence of external factors: If sharing was fully explained by internal determinants, then subjects should not make use of the option to sort out. (If anything, subjects who share zero might be indifferent.)

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 internal motivation. However, by the same logic, negative reciprocity should make people intrinsically less willing to share and therefore also less likely to avoid the sharing environment. Although this is not the case, we do observe variation in spiteful nonsharing (i.e., in the fraction of subjects who sort in but share nothing). Whereas only a small number of subjects sort in and share zero in the DG (11%), positive reciprocity reduces this rate further to 3%, and negative reciprocity increases the rate to 20%. In other words, nonsharers are not afraid to sort in and reveal their nonsharing decision (anonymously) to a partner who did not share him- or herself.[15] The reduction of the opt-out frequency under positive reciprocity and the increase in spiteful nonsharing under positive reciprocity point to variation in internal motives, but the effect is small and insignificant, confirming that external motives are also at play.

4. DISCUSSION

What are the implications of these results for our understanding of reciprocity and for theories that describe it? Existing theories have to wrestle with two key findings. First, in all three conditions, the sorting option has a significant impact on the distribution of amounts shared, and the impact is approximately invariant to reciprocity. Second, under positive reciprocity, we observe less opting

---

[15]More detailed analyses are given in **Supplemental Appendix B**. Whether such behavior truly reflects spite, as suggested in the reciprocity literature, might be debatable.

out and less spiteful nonsharing than in the DG, and under negative reciprocity we observe more of both. The increase in spiteful nonsharing under negative reciprocity is economically large.

We briefly recap why existing models fail to predict these findings. The main point, in all cases, is that none of these models predicts an impact of the mere sorting option on the amounts shared. People might sort out because they are not planning to share anyhow. But the sum of those opting out and those sorting in and sharing zero should be unaffected.

First, consider outcome-based models, such as in Fehr & Schmidt (1999). It is easy to see that these models successfully predict our finding that giving shifts downward under negative reciprocity but do not necessarily capture the increase in giving under positive reciprocity: In the DG, types who put sufficiently high weight on fairness or others' consumption share a positive portion of the $10. Everyone else is indifferent between sorting out or sorting in but sharing nothing. In the DDG, however, the distribution of the additional $2 might induce the dictator to share more or less to achieve his most preferred distribution of wealth.[16] Under negative reciprocity, 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. Under positive reciprocity, both the dictator and recipient have a head start of $1. Depending on the type of outcome-based model, the threshold level of fairness for sharing could go up (if dictators no longer feel the need to ensure a small minimum payment for the recipient), down (if the marginal utility of the eleventh dollar is particularly small), or stay the same (if dictators are merely inequity averse). 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. Hence, outcome-based models do not predict any effect of sorting nor the disproportionate increase in spiteful nonsharing.

Next, consider type-based models, such as Levine (1998). In these models, the dictator's altruism toward the recipient depends on his information about the recipient's type. In the standard game, the dictator has no information about his partner and chooses based on his knowledge about the population distribution of types. In the DDG, however, the dictator can draw inferences about his partner: He will have more favorable beliefs about the recipient after receiving $1, and less favorable ones after receiving $0.[17] As in the outcome-based models, however, the directional prediction under positive reciprocity remains ambiguous: Dictators may be less likely to share than in the DG condition, despite putting a higher weight on the recipient's consumptions, because the recipient is already starting out with $1. Moreover, similar to the outcome-based models, type-based models do not predict an effect of sorting. Any changes in giving should be driven by a shifting distribution of ideal allocations. Hence, the results about differential effects of the sorting option along with the rates of spiteful nonsharing cannot be explained by type-based reciprocity.

Finally, consider intentions-based models. As the recipient does not know about the later DG when she makes her mini-DG choice, we can analyze behavior as two normal form games and apply the model in Rabin (1993).[18] In the DG, the recipient has taken no action and is neutral. The dictator should therefore maximize his material outcome. In the DDG, the recipient's kindness in the mini-DG is negative if she shares nothing and positive if she shares $1. In the negative case, the

---

[16]In this section, labels of players always refer to their role in the primary (second-stage) DG. For example, the decision maker in the first-stage mini-DG is referred to as the recipient. Throughout, male pronouns are used for dictators, and female pronouns for recipients.

[17]Because the recipient was not informed about the main DG prior to making her choice in the minigame, we assume that this choice cannot be interpreted as a strategic signal.

[18]Technically, our game is an extensive-form game. Moreover, 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.

dictator simply maximizes his material outcome, as there is no avenue for costly punishment of the recipient. In the positive case, dictators with a high-enough weight on reciprocating intentions share a positive amount. Note that, in contrast to the models discussed above, this model makes unambiguous predictions for the positive reciprocity condition as players have utility not over their partners' outcomes but over each others' reciprocal intentions. However, once again, the sorting option should have no impact on the distribution of positive gifts in any condition. Hence, intentions-based reciprocity does not address the other conflicts with the data any better than outcome- or type-based models do.

To provide a bigger-picture intuition, we note that a sorting option makes the information acquired by the recipient 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. Anybody who modifies his behavior when given a sorting option must do so in response to this additional choice over information. The inadequacy of most reciprocity models to explain sorting behavior reflects their inability to deal with responses to observability.

So what can explain the experimental findings? Above we argue that the missing piece in existing models is external motivation. Individuals may share not only because they care about the other person's payoff or intentions, but because they feel obliged or pressured to share and would rather avoid being in this situation. There is a variety of terminology describing such external motivations, including social pressure, social norms, social image, social signaling, audience effects, prestige, shame, guilt, and reputation. We sometimes refer to social pressure or social image, but our analysis does not pin down the exact form of external motivation. Rather, it illustrates that a model of reciprocity that incorporates external factors can predict and provide an interpretation for the above results.

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, internal factors (such as fairness), and external factors (such as 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 no disutility from being subjected to social pressure to share. 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 social pressure induces him to share more than internally desired, and the benefit from not feeling such pressure, plus the material benefit from keeping the additional money, outweighs the costs of sharing less than internally desired. 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.

In the DDG, the role of external factors might change. Under positive reciprocity, external factors might become less important as positive reciprocity strengthens the internal motivation to give. Alternatively, positive reciprocity might increase the external pressure to share and make it less acceptable to share little, driving more people to sort out. Yet another possibility is that positive reciprocity strengthens external image rewards for sharing, and people might not opt out but share more. Under negative reciprocity, less sharing might reflect not only weaker internal motives to share, but also lower external motives. For example, a dictator who would share a positive amount in the no-sorting setting might now believe it to be fair to share less or feels less pressure to share. However, only external factors such as social image concerns predict, in a direct manner, a disproportionate 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 external determinants of giving may provide an interpretation for the reduced-form findings, including changes in donations and sorting-in rates that are not

positively correlated. The structural estimations will shed light on the relative strengths of the underlying motives and their interactions with reciprocity than we can glean from the reduced-form analysis.

5. MODEL AND ESTIMATION OF EXTRINSIC MOTIVES

We present a simple model of reciprocity that incorporates both internal and external determinants of sharing. Many alternative models of social pressure and other external 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 and nonspecific model of external determinants suffices to illustrate the underlying mechanism. We confirm this insight 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 internal 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 the subscript $r \in \{DG, NR, PR\}$ indicates the reciprocity environment, G_r is an increasing function of x, up to some maximal level (endowment), and $\alpha_r \in \mathbb{R}$ is the weight assigned to sharing. Thus, $-\alpha_r G_r$ captures internal motives to sharing as disutility from keeping too much. Note that both the disutility of keeping endowment and the weight may vary depending on the reciprocity environment. For simplicity, we assume that nonsharers, who are indifferent between sorting out and sorting in (and keeping all the money), sort out with some fixed probability that does not depend on the reciprocity context.

A simple way to incorporate external 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's actions are observable to the recipient, i.e., if he sorts in:

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

To reduce the degrees of freedom and allow for model identification, we impose the following specification:

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

A few details deserve comment. First, the quadratic loss function implies that people want to be generous but not too generous.[19] The specific 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. (In Section 5.2, we show that the specification is not critical to our findings.)

Second, the loss function is symmetric around \$5, which implies that we do not incorporate the initial \$2 mini-DG into the decision in the DDG. If we did, the loss function might still be centered

---

[19]This view is supported by, for example, Andreoni & Bernheim (2009). The debate whether overgenerosity induces guilt and a bad social image is still active (see Krupka & Weber 2013) but is not relevant to our estimations.

around \$5 in the positive reciprocity condition but around \$4 in the negative reciprocity condition. However, in our data, we continue to observe a few people sharing exactly \$5 in the negative reciprocity condition and no comparable bump in the distribution at \$4 (see **Figure 1c**).

Third, the external motivation is framed as a loss $[-\beta_r \mathbb{1}(\text{sort in})H_r(x)]$, which implies that, as long as β_r is positive, a giver cannot gain utility from sharing. At best, he obtains no loss in utility, namely when sharing the fair amount of \$5.[20] 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 α_r and β_r additively implies that internal and external determinants of sharing operate in parallel, with one exception: The two parameters may have opposite sign. An agent may feel negative altruism (spite) despite societal pressure to share or may like to share despite pressure to punish.

The sharing implications of this model are straightforward. In the absence of social pressure or other external factors, dictators keep the amount $x^* = 5 + 1/(2\alpha_r)$ (rounded to the nearest available discrete choice).[21] In the presence of external factors, they keep the adjusted amount $x^s = 5 + 1/(2(\alpha_r + \beta_r))$.[22] If the utility obtained with this adjusted choice is smaller than the utility when selfishly opting out and sharing zero, dictators sort out if possible. Higher internal motivation to share increases the cost of sorting out, and higher external pressure increases the cost of sharing moderately.

Figure 2 illustrates how the two type parameters, α_r and β_r, determine choices when sorting is allowed. A dictator with a $\beta_r < 0$ never chooses to opt out. That is, if selfishness induces a good social image, there is no benefit to sorting out rather than visibly sharing nothing. The more negative a dictator's β_r is, the higher must be his α_r in order for him to be persuaded to share anything at all. Among agents with $\beta_r > 0$, instead, higher levels of α_r still lead to more generosity among those who sort in, and higher levels of β_r force relatively ungenerous people to share more generously. But those who want to give small amounts would rather sort out and experience (internal) disutility from undersharing 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. Among dictators who have a low β_r, introducing a sorting option mostly affects the lowest givers, so the impact on average generosity should be small. Among dictators with a high β_r, generous givers will also opt out, namely, if they have low α_r's. Conversely, givers with low α_r's react to the introduction of a sorting option, and the opting out of the most generous givers among them (those with high β_r) will have a large impact, whereas givers with higher levels of α_r are more immune to the sorting option overall.

The figure also illustrates why existing models of reciprocity fail to explain the aspects of our data discussed in Section 3.2. Looking at the $\beta_r = 0$ axis, the only individuals who sort out are those who are also willing to sort in and share nothing ($\alpha_r \leq 0$). Hence, the systematic decrease in sorting after positive reciprocity and increase after negative reciprocity are not predicted.

---

[20]And, vice versa, if $\beta < 0$, there is a reward for punishing the other person, and sharing any amount leads to an increase in utility. This is almost certainly not true. The arguments apply also to the internal motivation for sharing.

[21]Note that x^* is a maximum iff $\alpha_r > 0$. If $\alpha_r < 0$, then x^* is a minimum, and we have a corner solution.

[22]Note that x^s is a maximum iff $\alpha_r + \beta_r > 0$. Otherwise, we have a corner solution. This becomes pertinent in the negative reciprocity case, in which we estimate that $\alpha_r + \beta_r < 0$.
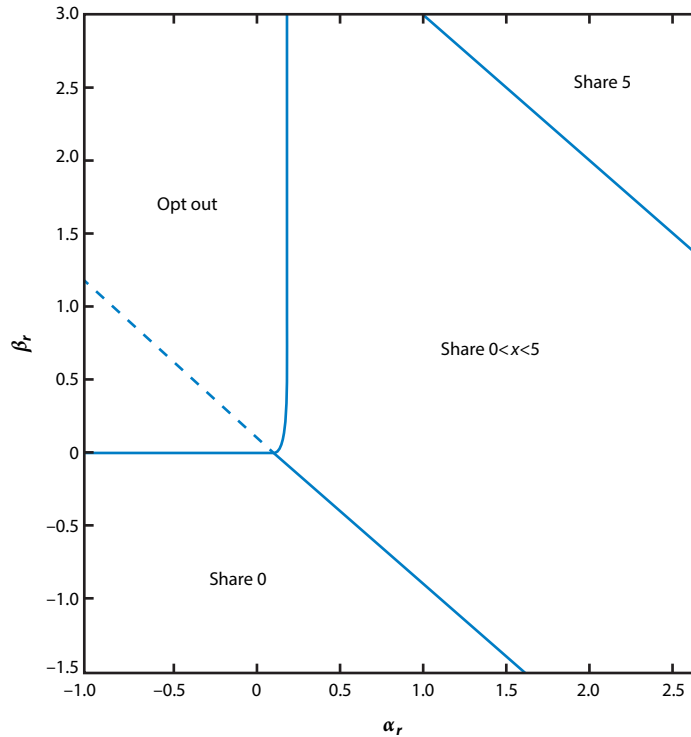
Figure 2

Parameterization of a reciprocity model.

Additionally, a mere shift in intrinsic motivation cannot explain a change in the distribution of gifts as, again, the only people willing to sort out are those who would share nothing in any case.

Finally, another way to look at **Figure 2** is to consider the possibility that individuals have the same β in all reciprocity environments, $\beta_{DG} = \beta_{PR} = \beta_{NR}$, whereas their α_r's may vary in r, for example, $\alpha_{NR} < \alpha_{DG} < \alpha_{PR}$. The figure shows that, as α_r increases, 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 and thereby induced a higher α. **Table 2** summarizes the predictions of three possible variants of the model in terms of the patterns we observe in the data. All variants allow for internal 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 abuse notation slightly and consider the type of reciprocity induced by an experimental condition, r, to be a continuous variable, with a more positive value denoting a more positive reciprocity context, i.e., kinder prior treatment. The three reciprocity treatments of the experiment, $r \in \{DG, NR, PR\}$, thus represent three choices of r from this continuous possibility set. In much of the following discussion, as convenient, we drop the r subscript on α and β with the understanding that they refer to individuals' type parameters within a particular reciprocity treatment and focus on the distributions of these parameters.

The first variant, shown in the first row, does not allow for external motivation ($\beta_\rho = 0$). This variant captures the existing outcome-, type-, and intentions-based models discussed above. The second variant, shown in the second row, allows for external motivation but assumes that it does

Table 2 Model predictions and data

	Gift distributions sorting dependent[a]	Sorting decreasing in r[b]	Spiteful nonsharing decreasing in r[c]	Impact of sorting on giving decreasing in r[d]	Average giving increasing in r[e]
No extrinsic motives: $\partial\alpha_r/\partial r > 0$, $\beta_r = 0$	No	Yes	No	No	Yes
Constant extrinsic motives: $\partial\alpha_r/\partial r > 0$, $\beta \sim F_\beta \forall r$	Yes	Yes	No	Yes	Yes
Variable extrinsic motives: $\partial\alpha_r/\partial r > 0$, $\partial\beta_r/\partial r > 0$	Yes	Possible	Possible	Possible	Possible
Data	Yes	Yes	Yes	Maybe	Yes

Comparison of three restrictions on α and β in the model of Section 5.1. 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 that the model is not able to predict that aspect of the data. Maybe indicates that the data produce a small but statistically insignificant effect consistent with the prediction. We assume here that α and β have full support so that even if with some distribution of types, most people do not change behavior when r changes; as long as someone does, the model is still said to predict a change overall.
[a]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 $F_{x,r,s} \neq F_{x,r,NS}$.
[b]Sorting is decreasing in r: $\partial P[U(10|\text{sort out}) > U(x^s|\text{sort in})]/\partial r < 0$.
[c]Spiteful nonsharing is decreasing in r: $\partial P[U(10|\text{sort in}) > U(x^s|\text{sort out})]/\partial r < 0$. Note that the term spiteful indicates a strict preference for sorting in and sharing nothing; equivalently, any increase in sorting in and sharing nothing is disproportionately more than an increase in sharing nothing overall.
[d]The impact of sorting on giving is decreasing in r: $\partial(\overline{x}_{r,S} - \overline{x}_{r,NS})/\partial r < 0$.
[e]Average giving is increasing in r: $\partial\overline{x}/\partial r < 0$.

not vary by prior kind treatment, or reciprocity condition, r. In the third variant, shown in the third row, external motivation is increasing in prior kindness. The last row shows the data.

The predictions for the effect of sorting, shown in the first column, provide the key reason to rule out existing models that do not account for external factors. It immediately follows that existing models do not capture variations in sorting due to variations in reciprocity (column 5). Moreover, column 3 shows that the disproportionate increase in spiteful non-sharers in the negative reciprocity setting is the main reason why an external 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 r and the relative magnitudes of α_ρ and β_ρ (see the entries "possible" in third row and "maybe" in the fourth row).

Our structural estimation will paint a clearer picture of the impact of reciprocity on internal and external 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 α_ρ and β_ρ for each environment from the data.

5.2. Estimation

We assume that α_r is normally distributed according to $N(\mu_{\alpha_r}, \sigma_{\alpha_r})$ and β_r is similarly distributed according to $N(\mu_{\beta_r}, \sigma_{\beta_r})$. We estimate the parameters $\mu_{\alpha_r}, \sigma_{\alpha_r}, \mu_{\beta_r}$, and σ_{β_r} for each $r \in \{DG, NR, PR\}$, using minimum distance estimation. For the vector of moments in the baseline estimation, we break down the choices of giving into bins: exactly \$0, from \$0.25 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. We have also estimated more constrained models, such as specifications that require the variances to be equal. Details on the estimation procedure and results are in **Supplemental Appendix C**.

An individual, i, with type parameters α_i and β_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 $U(x^s) > 10 - 25\alpha_i$ (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.

Column 1 of **Table 3** shows the results. We estimate μ_α in the DG to be significantly negative.[23] The estimate indicates that a majority of people do not like to share at all. Although our model is simplistic, this finding generally agrees with our and others' experimental findings that altruism toward strangers is widespread but far from universal. This baseline estimate implies that 31% of people would share a positive amount in a completely anonymous DG 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 (e.g., Koch & Normann 2008).

What is striking is the magnitude of β 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 nonreciprocal (one-sided) giving is perhaps more driven by external considerations such as social pressure than by internal social preferences.

The picture changes with the introduction of reciprocity. When the recipient has previously behaved selfishly (negative reciprocity treatment), internal altruism parameters plummet to the point at which only 16% of second-stage dictators feel any drive to share. At the same time, the weight of external 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.[24]

In the positive reciprocity environment, however, the weight of external factors remains very similar to that 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.

Next, we contrast our estimates with those based on models that do not allow the external motivation to depend on the reciprocity environment (constant extrinsic motivation) and with estimates based on standard reciprocity models (no extrinsic motivation). Column 2 of **Table 3** shows the estimates when external motivation is held constant across reciprocity environments, and column 3 shows the estimates when external motivation is required to be zero. 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 nonsharers sort out with some fixed probability), this specification fits the data best and creates the most conservative comparison to the model with external factors.

---

[23]Because we do not investigate outright spite, we cannot distinguish this result from an alternative specification in which altruism is censored at zero; the key prediction is that a majority of people do not put positive weight on others' outcomes.

[24]External determinants, such as social pressure and social norms, are usually thought of as attributes of a situation that apply equally to everyone, but because 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.

Table 3 Structural model comparison

	Estimated parameter	Variable extrinsic motive	Constant extrinsic motive	No extrinsic motive
Dictator game	μ_α	−1.732 (0.486)	−1.643 (0.244)	−0.294 (0.452)
	σ_α	3.569 (0.757)	3.395 (0.347)	6.031 (0.796)
	μ_β	2.560 (0.489)	2.435 (0.193)	
	σ_β	3.159 (0.725)	3.372 (0.239)	
Negative reciprocity	μ_α	−5.723 (2.073)	−7.561 (0.453)	−3.938 (1.064)
	σ_α	5.789 (2.124)	6.922 (0.695)	6.920 (1.423)
	μ_β	1.010 (0.657)	2.435 (0.193)	
	σ_β	1.627 (1.071)	3.372 (0.239)	
Positive reciprocity	μ_α	0.153 (0.140)	0.123 (0.211)	1.389 (0.352)
	σ_α	1.247 (1.005)	1.961 (0.276)	5.220 (0.557)
	μ_β	2.893 (0.450)	2.435 (0.193)	
	σ_β	3.485 (0.489)	3.372 (0.239)	
Weighted SSE		288.671	291.839	380.243

Generalized method of moments estimation results for baseline model specification (33 moments, α and β normally distributed) and comparison models requiring that social pressure be invariant to reciprocity, or zero social pressure. α refers to the weight on internal altruism and β to the weight on external social image. SSE is the sum of squared errors of the estimated moments, and parentheses denote standard errors.

Nonetheless, a model that does not allow for social pressure or other external factors does significantly worse at fitting the data—as shown in column 3, the weighted sum of squared errors is much higher (380 versus 289). 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 external motives but restricts them to be reciprocity invariant instead yields estimates that are rather similar to those we obtain when we allow external motivation to vary. The estimates for the constant-β 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 internal 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 β. 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 internal but also external factors does not significantly improve the fit of the model, relative to assuming reciprocity-invariant external factors. This result parallels the reduced-form results that the impact of sorting on giving is approximately invariant across reciprocity environments.

We conclude that although reciprocity may slightly influence external motivations to share, it primarily acts through the channel of internal motives. Incorporating reciprocity-invariant external pressure into theories of reciprocity greatly improves their predictive power, but an additional interaction between external factors and reciprocity has a relatively small impact.

At the same time, the estimation results also show that the amount of additional internal motivation—in the sense of reciprocity-induced altruism or similar heightened internal motives

under positive reciprocity—is estimated approximately correctly under any of the models, including the naïve model that does not allow for social pressure or other extrinsic motives. Specifically, if one asks how much additional internal willingness to share is induced by a kind treatment of the recipient, the naïve model implies that the average α_{NR} exceeds the average α_{DG} by 1.683. Under the models that allow for external motivation, the corresponding differences between the μ_α's equal 1.885 (variable β) and 1.766 (constant β) and are hence quite similar. This robustness reflects that external motives are remarkably stable across both environments.

This conclusion is somewhat less true for the negative reciprocity environment: The decrease in internal willingness to share after unkind treatment, relative to a neutral treatment, is slightly underestimated when neglecting external factors (the difference in μ_α's amounts to -0.347) and is strongly overestimated when allowing only for a reciprocity-insensitive external motive (with a difference of 1.927), relative to the model with reciprocity-dependent β's. 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.

6. CONCLUSION

This article questions whether reciprocal behavior primarily reflects internal motivations to share. Recent research highlights the importance of incorporating external factors, such as social pressure and a concern for social image, into models of prosocial behavior. We provide an overview of this second wave of research on social preferences. We then point out that this more recent research has so far had little impact on how economists view and model reciprocity.

To address this potential gap in the literature, we use a novel set of experimental conditions and show that reciprocal behavior responds to avoidance opportunities in ways unaccounted for by existing models of reciprocity. The experimental conditions compare sharing behavior in social environments in which positive (or negative) reciprocity is in play, while varying whether 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 (DG) environment. We also find that the rate of spitefully sorting in without sharing anything decreases in the inferred kindness of the recipient (i.e., out of proportion to the general rate of sharing zero). Traditional models of internally 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, external motivations are required to explain behavior.

We also show how one might incorporate external determinants (e.g., social pressure) into a model that is similar in spirit to previous reciprocity models and 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 external motivations strongly improves the predictive power of the model. We also confirm, however, that reciprocity primarily changes behavior by changing people's internal motivation to share, consistent with existing models.

In terms of future research, further experiments that distinguish internal and external motivations for prosocial behavior are needed. Similarly, the literature would strongly benefit from field experiments exploring whether reciprocity in real-world economic settings, such as contracts and labor markets, may also reflect external motives. For example, do above-market wages induce more effort due to increased altruism toward the employer, or might they also reflect social pressure to work hard when paid well?

As another important and bigger step, research needs to pin down the type of preferences underlying our general term external motivation. For example, to what extent are external

determinants of sharing social image concerns, and to what extent do they reflect direct social pressure? Whereas social pressure would imply that reciprocal giving reduces welfare, the net welfare effect of social image concerns is less clear. Is there an image reward for reciprocity? 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 external motivation.

Finally, internal motives for reciprocity need rethinking beyond the confounds with external factors. Even motives that are cleanly identified as internal—e.g., because the giver seeks rather than avoids an (anonymous, invisible) giving opportunity—are not necessarily altruism or fairness. Instead, a giver might feel an internal obligation to reciprocate in the sense of Sugden (1984) and Cialdini (1993) and as proposed in the anthropology and sociology literature. For example, Malmendier & 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 intentions-based or type-based view of reciprocity but consistent with an obligation- or norm-based understanding (see also Bicchieri 2006, Postlewaite 2011, Krupka et al. 2012). This distinction is particularly relevant for our understanding of the role of reciprocity in markets. When a firm gives gifts with the intention to maximize profits, by inducing reciprocity among customers, 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.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

Akerlof GA. 1982. Labor contracts as partial gift exchange. *Q. J. Econ.* 97:543–69

Akerlof GA, Kranton RE. 2000. Economics and identity. *Q. J. Econ.* 115:715–53

Akerlof GA, Yellen JL. 1988. Fairness and unemployment. *Am. Econ. Rev.* 78:44–49

Akerlof GA, Yellen JL. 1990. The fair wage-effort hypothesis and unemployment. *Q. J. Econ.* 105:255–83

Alpizar F, Carlsson F, Johansson-Stenman O. 2008. Does context matter more for hypothetical than for actual contributions? Evidence from a natural field experiment. *Exp. Econ.* 11:299–314

Andreoni J. 1989. Giving with impure altruism: applications to charity and Ricardian equivalence. *J. Polit. Econ.* 97:1447–58

Andreoni J. 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ. J.* 100:464–77

Andreoni J, Bernheim BD. 2009. Social image and the 50-50 norm: a theoretical and experimental analysis of audience effects. *Econometrica* 77:1607–36

Andreoni J, Miller JH. 2002. Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica* 70:737–53

Andreoni J, Vesterlund L. 2001. Which is the fair sex? Gender differences in altruism. *Q. J. Econ.* 116:293–312

Battigalli P, Dufwenberg M. 2007. Guilt in games. *Am. Econ. Rev.* 97:170–76

Battigalli P, Dufwenberg M. 2009. Dynamic psychological games. *J. Econ. Theory* 144:1–35

Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD. 2001. Bad is stronger than good. *Rev. Gen. Psychol.* 5:323–70

Becker GS. 1961. *Notes on an economic analysis of philanthropy*. Work. Pap., Natl. Bur. Econ. Res.

Bénabou R, Tirole J. 2006. Incentives and prosocial behavior. *Am. Econ. Rev.* 96:1652–78

Berg J, Dickhaut J, McCabe KA. 1995. Trust, reciprocity, and social history. *Games Econ. Behav.* 10:122–42

Bergstrom T, Blume L, Varian H. 1986. On the private provision of public goods. *J. Public Econ.* 29:25–49

Bewley TF. 2009. *Why Wages Don't Fall During a Recession*. Cambridge, MA: Harvard Univ. Press

Bicchieri C. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge, UK: Cambridge Univ. Press

Blount S. 1995. When social outcomes aren't fair: the effect of causal attributions on preferences. *Organ. Behav. Hum. Decis. Process.* 63:131–44

Bohnet I, Frey BS. 1999a. Social distance and other-regarding behavior in dictator games: comment. *Am. Econ. Rev.* 89:335–39

Bohnet I, Frey BS. 1999b. The sound of silence in prisoner's dilemma and dictator games. *J. Econ. Behav. Organ.* 38:43–57

Bolton GE, Ockenfels A. 2000. ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90:166–93

Bowles S, Gintis H. 2011. *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton, NJ: Princeton Univ. Press

Brandts J, Solà C. 2001. Reference points and negative reciprocity in simple sequential games. *Games Econ. Behav.* 36:138–57

Brennan TA, Rothman DJ, Blank L, Blumenthal D, Chimonas SC, et al. 2006. Health industry practices that create conflicts of interest. *JAMA* 295:429–33

Broberg T, Ellingsen T, Johannesson M. 2007. Is generosity involuntary? *Econ. Lett.* 94:32–37

Brock JM, Lange A, Ozbay EY. 2013. Dictating the risk: experimental evidence on giving in risky environments. *Am. Econ. Rev.* 103:415–37

Camerer CF, Thaler RH. 1995. Anomalies: ultimatums, dictators and manners. *J. Econ. Perspect.* 9(2):209–19

Cappelen AW, Hole AD, Sørensen EØ, Tungodden B. 2007. The pluralism of fairness ideals: an experimental approach. *Am. Econ. Rev.* 97:818–27

Carpenter JP, Matthews PH. 2004. *Social reciprocity*. Discuss. Pap. 1347, IZA, Bonn

Carpenter JP, Matthews PH, Ongonga O. 2004. Why punish? Social reciprocity and the enforcement of prosocial norms. *J. Evol. Econ.* 14:407–29

Charness G, Rabin M. 2002. Understanding social preferences with simple tests. *Q. J. Econ.* 117:817–69

Cherry TL, Frykblom P, Shogren JF. 2002. Hardnose the dictator. *Am. Econ. Rev.* 92:1218–21

Cialdini RB. 1993. *Influence: The Psychology of Persuasion*. New York: HarperCollins

Cooper DJ, Kagel JH. 2014. Other-regarding preferences: a selective survey of experimental results. In *Handbook of Experimental Economics*, Vol. 2, ed. JH Kagel, AE Roth. Princeton, NJ: Princeton Univ. Press. In press

Cox JC. 2004. How to identify trust and reciprocity. *Games Econ. Behav.* 46:260–81

Cox JC, Friedman D, Gjerstad S. 2007. A tractable model of reciprocity and fairness. *Games Econ. Behav.* 59:17–45

Croson RTA. 2007. Theories of commitment, altruism and reciprocity: evidence from linear public goods games. *Econ. Inq.* 45:199–216

Dana J, Cain DM, Dawes RM. 2006. What you don't know won't hurt me: costly (but quiet) exit in dictator games. *Organ. Behav. Hum. Decis. Process.* 100:193–201

Dana J, Weber RA, Kuang JX. 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Econ. Theory* 33:67–80

DellaVigna S, List JA, Malmendier U. 2012. Testing for altruism and social pressure in charitable giving. *Q. J. Econ.* 127:1–56

DellaVigna S, List JA, Malmendier U, Rao G. 2013. The importance of being marginal: gender differences in generosity. *Am. Econ. Rev.* 103:586–90

Dhaene G, Bouckaert J. 2010. Sequential reciprocity in two-player, two-stage games: an experimental analysis. *Games Econ. Behav.* 70:289–303

Dufwenberg M, Kirchsteiger G. 2004. A theory of sequential reciprocity. *Games Econ. Behav.* 47:268–98

Falk A. 2007. Gift exchange in the field. *Econometrica* 75:1501–11

Falk A, Fischbacher U. 2006. A theory of reciprocity. *Games Econ. Behav.* 54:293–315

Falk A, Fehr E, Fischbacher U. 2008. Testing theories of fairness: Intentions matter. *Games Econ. Behav.* 62:287–303

Fehr E, Fischbacher U. 2004. Third-party punishment and social norms. *Evol. Hum. Behav.* 25:63–87

Fehr E, Gächter S. 2000a. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* 151:867–68

Fehr E, Gächter S. 2000b. Fairness and retaliation: the economics of reciprocity. *J. Econ. Perspect.* 14:159–82

Fehr E, Gächter S, Kirchsteiger G. 1997. Reciprocity as a contract enforcement device: experimental evidence. *Econometrica* 65:833–60

Fehr E, Goette L, Zehnder C. 2009. A behavioral account of the labor market: the role of fairness concerns. *Annu. Rev. Econ.* 1:355–84

Fehr E, Kirchsteiger G, Riedl A. 1993. Does fairness prevent market clearing? An experimental investigation. *Q. J. Econ.* 108:437–59

Fehr E, Schmidt KM. 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114:817–68

Feld LP, Frey BS. 2007. Tax compliance as the result of a psychological tax contract: the role of incentives and responsive regulation. *Law Policy* 29:102–20

Finan F, Schechter L. 2012. Vote-buying and reciprocity. *Econometrica* 80:863–81

Fong CM, Luttmer EFP. 2009. What determines giving to Hurricane Katrina victims? Experimental evidence on racial group loyalty. *Am. Econ. J. Appl. Econ.* 1:64–87

Franzen A, Pointner S. 2012. Anonymity in the dictator game revisited. *J. Econ. Behav. Organ.* 81:74–81

Gale J, Binmore KG, Samuelson L. 1995. Learning to be imperfect: the ultimatum game. *Games Econ. Behav.* 8:56–90

Geanakoplos J, Pearce D, Stacchetti E. 1989. Psychological games and sequential rationality. *Games Econ. Behav.* 1:60–79

Gerber AS, Green DP, Larimer CW. 2008. Social pressure and voter turnout: evidence from a large-scale field experiment. *Am. Polit. Sci. Rev.* 102:33–48

Gintis H. 2000. Strong reciprocity and human sociality. *J. Theor. Biol.* 206:169–79

Gneezy U, List JA. 2006. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. *Econometrica* 74:1365–84

Greenberg J. 1990. Employee theft as a reaction to underpayment inequity: the hidden cost of pay cuts. *J. Appl. Psychol.* 75:561–68

Grossman Z. 2012. *Self-signaling and social-signaling in giving*. Work. Pap., Dep. Econ., Univ. Calif., Santa Barbara.

Harbaugh WT. 1998. What do donations buy? A model of philanthropy based on prestige and warm glow. *J. Public Econ.* 67:269–84

Henrich J, Boyd R, Bowles S, Camerer CF, Fehr E, et al. 2001. In search of *Homo economicus*: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* 91:73–78

Henrich J, Ensminger J, McElreath R, Barr A, Barrett C, et al. 2010. Markets, religion, community size, and the evolution of fairness and punishment. *Science* 327:1480–84

Hoffman E, McCabe KA, Shachat K, Smith VL. 1994. Preferences, property rights, and anonymity in bargaining games. *Games Econ. Behav.* 7:346–80

Hoffman E, McCabe KA, Smith VL. 1996. Social distance and other-regarding behavior in dictator games. *Am. Econ. Rev.* 86:653–60

Hoffman E, McCabe KA, Smith VL. 1998. Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. *Econ. Inq.* 36:335–52

Johannesson M. 2000. Non-reciprocal altruism in dictator games. *Econ. Lett.* 69:137–42

Keser C, van Winden F. 2000. Conditional cooperation and voluntary contributions to public goods. *Scand. J. Econ.* 102:23–39

Koch AK, Normann HT. 2008. Giving in dictator games: regard for others or regard by others? *South. Econ. J.* 75:223–31

Krueger AB, Mas A. 2004. Strikes, scabs, and tread separations: labor strife and the production of defective Bridgestone/Firestone tires. *J. Polit. Econ.* 112:253–89

Krupka EL, Leider S, Jiang M. 2012. *A meeting of the minds: contracts and social norms.* Work. Pap., Univ. Mich., Ann Arbor

Krupka EL, Weber RA. 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? *J. Eur. Econ. Assoc.* 11:495–524

Kube S, Maréchal MA, Puppe C. 2012. The currency of reciprocity: gift exchange in the workplace. *Am. Econ. Rev.* 102:1644–62

Kube S, Maréchal MA, Puppe C. 2013. Do wage cuts damage work morale? Evidence from a natural field experiment. *J. Eur. Econ. Assoc.* 11:853–70

Lazear EP, Malmendier U, Weber RA. 2012. Sorting in experiments with application to social preferences. *Am. Econ. J. Appl. Econ.* 4:136–64

Ledyard JO. 1997. Public goods: a survey of experimental research. In *The Handbook of Experimental Economics*, Vol. 1, ed. JH Kagel, AE Roth, pp. 111–94. Princeton, NJ: Princeton Univ. Press

Levine DK. 1998. Modeling altruism and spitefulness in experiments. *Rev. Econ. Dyn.* 1:593–622

List JA. 2006. The behavioralist meets the market: measuring social preferences and reputation effects in actual transactions. *J. Polit. Econ.* 114:1–37

MacKinnon JG, White H. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econom.* 29:305–25

Malmendier U, Schmidt KM. 2012. *You owe me.* NBER Work. Pap. 18543

Mas A, Moretti E. 2009. Peers at work. *Am. Econ. Rev.* 99:112–45

Okun AM. 1981. *Prices and Quantities: A Macroeconomic Analysis.* Washington, DC: Brookings Inst.

Ostrom E, Walker J, Gardner R. 1992. Covenants with and without a sword: Self-governance is possible. *Am. Polit. Sci. Rev.* 86:404–17

Pillutla MM, Malhotra D, Murnighan JK. 2003. Attributions of trust and the calculus of reciprocity. *J. Exp. Soc. Psychol.* 39:448–55

Postlewaite A. 2011. Social norms and social assets. *Annu. Rev. Econ.* 3:239–59

Rabin M. 1993. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83:1281–302

Rotemberg JJ. 2008. Minimally acceptable altruism and the ultimatum game. *J. Econ. Behav. Organ.* 66:457–76

Roth AE, Erev I. 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games Econ. Behav.* 8:164–212

Sobel J. 2005. Interdependent preferences and reciprocity. *J. Econ. Lit.* 151:867–68

Sugden R. 1984. Reciprocity: the supply of public goods through voluntary contributions. *Econ. J.* 94:772–87

van der Weele J, Kulisa J, Kosfeld M, Friebel G. 2014. Resisting moral wiggle room: How robust is reciprocity? *Am. Econ. J. Microecon.* In press