On The Role of Sacrifice in Reciprocity*

Simin He† and Jiabin Wu‡

December 10, 2022

Abstract

This paper experimentally investigates the importance of sacrifice in affecting people's reciprocal behavior. Our design allows us to exactly pin down how sacrifice of the sender's own payoff matters for her perceived kindness from the eyes of the receiver in a sender-receiver game, without being confounded by fairness concerns and higher order beliefs. We show that a simple extension of the axiomatic model of reciprocity by Cox et al. [2008] can nicely incorporate sacrifice, which matches the new empirical regularities found in our experiment.

Keywords: Sacrifice, Reciprocity, Intention, Kindness, Social Preferences, Experimental Economics.

JEL Codes: C92, D91.

Declarations of interest: none.

*The aurthors sincerely thank Michael Kuhn for his constructive comments that help to improve the paper. The authors are also grateful to Xintong Pan for her excellent research assistance, and to Tim Cason, Amanda Chuan, the participants of Warwick-SUFE workshop, WEA annual conference, ESA North American Meetings, XMU International Workshop on Experimental Economics for their helpful comments. Simin He acknowledges the National Natural Science Foundation of China (grant 72022010, 71803111) for its financial support.

†School of Economics, Shanghai University of Finance and Economics, 111 Wuchuan Rd, 200433 Shanghai, China. E-mail: he.simin@mail.shufe.edu.cn.

‡Department of Economics, University of Oregon, 515 PLC 1285 University of Oregon, Eugene, OR, USA 97403. E-mail: jwu5@uoregon.edu.

1 Introduction

Reciprocity, as an important type of social preferences, has been extensively studied by economists for decades. Güth et al. [1982] is the earliest experiment demonstrating the possible existence of reciprocity. They find that in the ultimatum game, the receiver tends to reject low offers from the sender even if it is costly to do so. Fehr et al. [1993] conduct the first experiment on the gift-exchange game, which supports the notion that workers tend to reciprocate firm's more generous offers by exerting higher efforts [Akerlof, 1982, Akerlof and Yellen, 1988, 1990]. These experimental works and many that follow (see Güth and Kocher [2014] for a survey on experimental ultimatum games and Charness and Kuhn [2011] for a survey on experimental gift-exchange games) spark the development of different theoretical models [Rabin, 1993, Levine, 1998, Charness and Rabin, 2002, Dufwenberg and Kirchsteiger, 2004, Falk and Fischbacher, 2006, Cox et al., 2007, 2008]. See Sobel [2005] and Battigalli and Dufwenberg [2021] for surveys.[1] Roughly speaking, all of them posit that one's desire to reward or punish others depends on whether those others have treated her kindly or not.

How does an individual determine whether they have been treated kindly? Many of the aforementioned works suggest that the intention behind an action plays a key role. However, intentions are rarely observed and instead must be inferred. For a given action, that inference could depend on its costs, benefits, or both. First, the perceived kindness of an action should increase in the magnitude the favor one does to that individual. We call this the principle of helping. Second, the perceived kindness of an action should also depend on how much one has to sacrifice to do the favor. We call this the principle of sacrifice. While different models differ in their definitions of intention, they all agree upon the principle of helping. Yet, to our surprise, the principle of sacrifice is largely ignored in the literature.[2] There are many daily life observations of the principle of sacrifice. When

[1] See also recent developments by Sebald [2010], Dufwenberg et al. [2013], Çelen et al. [2017], Jiang and Wu [2019], Sohn and Wu [2021].

[2] The notion of sacrifice is discussed in Falk and Fischbacher [2006] (see also Charness and Rabin [2002]). Yet the role of sacrifice they consider is more subtle than the principle of sacrifice we posit above, which we will discuss in details in Section 4. Brandts et al. [2014] also consider sacrifice in their experiment. But

one of your employees managed to finish an urgent report last night, how you may want to compensate the employee depends on whether you are aware that it was his or her spouse's birthday. When you wish to reciprocate a colleague who helped debug your code at work, it may matter whether your colleague was busy at the time. When a friend gives you a ride home after a party, you may view the favor differently if he or she had to make a detour. The principle of sacrifice may have a historical root in the hunter-gatherer societies. For example, a tribe in the Washington State, the Lummi Nation, tries to preserve a reciprocal norm between human and salmon through a story of the "Salmon Woman," who saved people from starvation by sacrificing her own children.[3]

This paper attempts to achieve two goals: 1) to find experimental supports for the principle of sacrifice; 2) to refine the current theory of reciprocity to incorporate sacrifice. It is not easy to use some standard experimental games to obtain a clean test of how sacrifice matters. To see why, consider the two simple sender-receiver games shown in Figure 1.

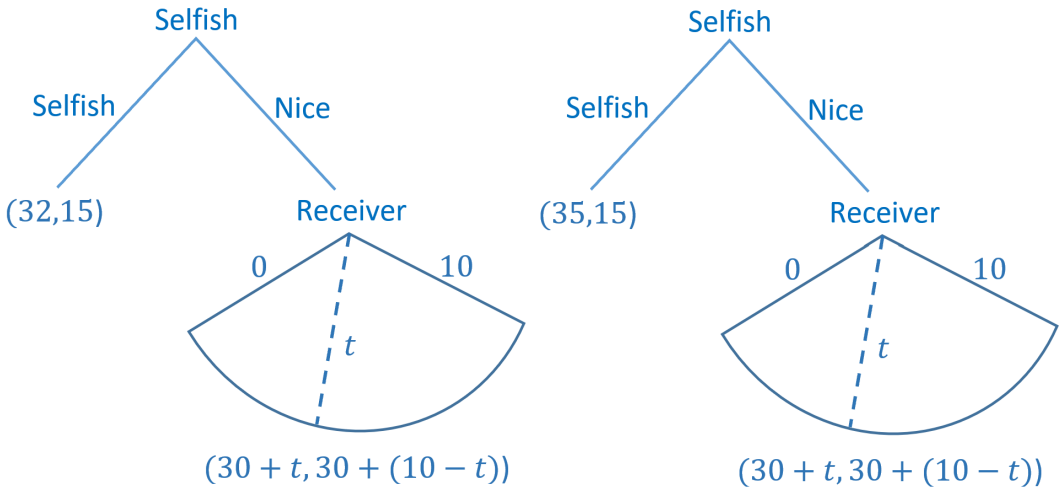

Figure 1: Two Sender-Receiver Games

they do not focus on the role of sacrifice in determining perceived intention involved in reciprocal decisions. McCabe et al. [2003] investigate the role of intention in positive reciprocity, by varying whether the senders voluntarily or passively forgo a positive opportunity cost to act cooperatively in a simple sender-receiver game. The paper naturally touches on the idea of sacrifice. However, it is hard to tell what exactly sacrifice is in their experiment.

[3]See https://nautil.us/reciprocity-in-the-age-of-extinction-11693/#! for the article about the convention.

In the first stage, the sender (he) can choose between two actions: *Selfish* and *Nice*. If the sender chooses the Selfish action, the game ends and he gets an outside option of either 32 or 35, while the receiver (she) gets 15. If he instead chooses the Nice action, both him and the receiver get 30 and the game proceeds to the second stage, in which the receiver can choose to allocate an additional 10 between herself and the sender. Let $t \in [0, 10]$ denote the amount allocated to the sender. Upon observing the sender choosing the Nice action, how does the receiver perceive the intention of the sender? Although, on the surface, it seems that the sender needs to sacrifice more to choose the Nice action when his outside option is 35, the receiver may believe that the sender expects her to allocate more to him at the second stage of the game. Hence, it is unclear whether the receiver actually think that the sender sacrifices more. The problem of using games like these two sender-receiver games is that one cannot control the receiver's belief about the sender's belief about her action (the sender's second order belief).

To overcome this problem, we employ a novel design. We use three sender-receiver games similar to those described in Figure 1 (with an additional one in which the sender's outside option is 28), but with three critical differences. First, we make the second stage of the games as a surprise round such that neither player knows its existence until the end of the first stage. Second, in the surprise round, we allow the receiver to allocate the additional 10 even when the sender has chosen Selfish in the experiment. Finally, we let the senders make choices in all three games and the outcome is determined by one randomly drawn game out of the three. The receiver sees the outcome and the full set of choices made by the sender. Then she is notified that she can choose how to allocate the 10. Given this design, the senders' choices are purely motivated by his other regarding preferences without expecting any favor in return from the receiver and the receiver is aware of this when making her choice of allocation. A sender's willingness to sacrifice is directly revealed by the choices he makes in all the three games, and the relationship between the receiver's allocation and the sender's choice profile —holding the outcome of the selected game fixed— measures the strength of the principle of sacrifice. The relationship between the receiver's allocation and the outcome they experience (15 or 30) —holding fixed the sender's choice profile— measures the strength of the principle

of helping. In addition, we control for fairness concern of the receiver by making the outcomes of both players identical across the three games whenever the sender chooses Nice.

We find clear support for both the principle of helping and the principle of sacrifice. The experimental results urge us to rethink about the existing theories. While higher order beliefs clearly matter for reciprocity as emphasized by Rabin [1993], Charness and Rabin [2002], Dufwenberg and Kirchsteiger [2004], Falk and Fischbacher [2006], we do not choose to work with their models because our focus is not on the role of higher order beliefs and our design frees us from considering them. Instead, we choose to extend the axiomatic model by Cox et al. [2008] because of its generality. We show that by adding a measure of the intensity of generosity, which is missing in Cox et al. [2008], and developing an axiom on how people react to different intensities of generosity can nicely accommodate the principle of sacrifice without compromising the principle of helping.

In a related work, Orhun [2018] studies how perceived motives affect reciprocity level, and she makes a clear distinction between perceived motives and perceived intentions. She finds that, when the same beneficial action of the first-movers is more likely to be motivated by strategic incentives rather than pure altruism, the receivers exhibit a lower level of positive reciprocity. See also Stanca et al. [2009] and Johnsen and Kvaløy [2016] for related discussions on the role of strategic motives in reciprocity. In contrast, our study makes a contribution to the intention-based reciprocity by enriching the measurement of intention, as we incorporate the dimension of sacrifice. Moreover, via our design, we achieve a clean test of the effect of sacrifice by excluding strategic motives of the first-movers.

The paper is organized as follows. Section 2 provides the details of the experimental design, procedures and establishes the hypotheses. Section 3 presents the experimental results. Section 4 discusses in depth the existing theories and provides the new theory. Section 5 concludes.

2 Experimental design, procedures and hypotheses

2.1 Design

In this experiment, we employ a simple binary dictator game. Subjects are randomly and anonymously matched in pairs to play the dictator game. One in each pair is randomly selected to be the sender, while the other becomes the sender. In this dictator game, the sender can choose between two options, Selfish and Nice (labelled in a neutral manner in the experiment). The game tree is depicted in Figure 2 below.

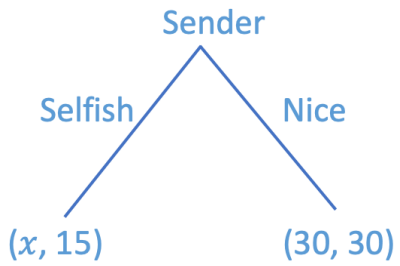

Figure 2: The experimental game

As can be seen from Figure 2, the Nice action always yields 30 points to both the sender (player 1, he) and the receiver (player 2, she), whereas the Selfish action always yields 15 points to the receiver. In the experiment, we vary the number x, which denotes the outside option of the sender. The number x takes the value of 28, 32 or 35. One can regard $x-30$ as the amount that the sender needs to sacrifice to help the receiver.

To measure the kindness of the sender, we employ a within-subject design with strategy method. All senders in the experiment have to make choices for all the three different dictator games with $x = 28, 32, 35$. The payoffs are determined by one randomly drawn game out of the three games. This design allows us to make two types of comparisons: (1) conditioned on the same type of senders (senders who make three identical choices), we can examine the effect of different outcomes on reciprocity, (2) conditioned on the same outcome (30, 30), we can look at how the level of reciprocity is affected by the type of the senders (senders who make different choices but end up with the same outcome).

During the experiment, after the senders make all the choices in the three dictator

games, one game is randomly drawn to determine the payoffs. Then, the complete choice profiles of the senders and the realized payoffs are revealed to the matched receivers. Here, we introduce a surprise round to elicit the positive reciprocity level of the receivers. In this surprise round, we provide a total of 10 points to the receivers, and ask the receivers to allocate these 10 points, in integer numbers only, between themselves and their matched sender. That is, these 10 points are provided in addition to the payoffs received in the dictator game. Note that, both the receivers and the senders are unaware of the surprise round. This is a critical part of our design. The reason of such a design is to control for higher order beliefs. If the senders are aware of the surprise round, in which the receivers can reward them in an allocation task, their choices in the dictator games can become strategic; and this will in turn affect the beliefs of the receivers and how they perceive the kindness of the senders. Our design of the surprise round effectively eliminates any higher order belief. Note also that the outcome after the sender chooses Nice is $(30, 30)$, which is identical across games. Hence, we control for the receiver's distributional concerns.

2.2 Procedures

The experiment was conducted at the Shanghai University of Finance and Economics in May-June 2021 and September 2022. Chinese subjects were recruited from the subjects pool of the Economic Lab via Ancademy.[4] We ran 15 sessions in total. Depending on the number of people showing up in the experiment, the number of subjects participated per session ranged from 8 to 30. In total, 328 subjects were recruited.[5]

The experiment was computerized using z-Tree [Fischbacher, 2007] and conducted in Chinese.[6] Upon arrival, subjects were randomly assigned a card indicating their table number and were seated in the corresponding cubicle. Instructions were displayed on their computer screens. Control questions were conducted to check their understanding of the instructions. The same experimenters were always presented during all the experimental sessions.

[4] Ancademy is a platform for social sciences experiments.

[5] Summary statistics of the subjects are provided in Table 3 in Appendix B.

[6] The English translations of instructions and screenshots are provided in Appendix A.

After finishing the experiment, subjects received their earnings privately through mobile payment.[7] Average earnings were ¥48 (equivalent to around 7 US dollars), including a show-up fee of ¥15 (around 2 US dollars). Each session lasted about 30 minutes.

2.3 Hypotheses

We hypothesize that the kindness of player 1, which depends on how much he sacrifices to choose the Nice action, and the outcome of player 2 both matter for the level of reciprocity of player 2.

To formulate our hypotheses, we first classify all the possible scenarios in our experiment. Based on our experimental design, there exist three different "player 1's types" and 2 possible "final payoffs" faced by player 2. In principle, there are only three types of sender that satisfy both individual utility maximization and the monotonicity of social preferences: a) Selfish: "choose Nice if $x = 28$, choose Selfish otherwise." These type-a senders are not willing to sacrifice their own payoffs to help the receivers at all. b) Somewhat nice: "choose Nice if $x = 28$ or 32, choose Selfish otherwise." The type-b senders are only willing to sacrifice up to 2 to help the receivers. c) Very nice: "choose Nice in all the possible values of x." The type-c senders are willing to sacrifice up to 5 to help the receivers.[8] When facing a type-a sender, the receiver may receive 15 or 30 ($a15$, $a30$). When facing a type-b sender, the receiver may receive 15 or 30 ($b15$, $b30$). Finally, when facing a type-c sender, the receiver will always receive 30 ($c30$). In sum, there are five possible combinations of the sender's type and the final payoff: $a15, a30, b15, b30, c30$.

Hypothesis 1. Conditioned on the same outcome, a kinder sender type induces a higher level of reciprocity of the receiver. That is, the levels of reciprocity satisfy $c30 > b30 > a30$, and $b15 > a15$.

Hypothesis 2. Conditioned on the same type of sender, a higher outcome of the receiver

[7] We used the payment system of Ancademy.

[8] While it is also possible for a sender to choose Selfish in all games, this behavior is rare in our study, and thereby we exclude those data points and do not formulate a corresponding hypothesis here.

induces a higher level of reciprocity of the receiver. That is, the levels of reciprocity satisfy $a30 > a15$, and $b30 > b15$.

3 Results

3.1 Types of senders

In total, we have 164 pairs of senders and receivers in our experiment. Among the 164 senders, there are in total four types of choice profiles: 73 senders (44.5%) choose Nice when $x = 28$, and choose Selfish otherwise (type-a, selfish), 28 senders (17.1%) choose Nice when $x = 28$ or 32, and choose Selfish if $x = 35$ (type-b, somewhat nice), and 58 (35.4%) senders choose Nice in all possible values of x (type-c, very nice). Finally, there are 5 senders (3.0%) who choose Selfish for all possible values of x (type-a', anti-social), we exclude them in the analysis because they are rare.

Because of our design of the surprise round, one may worry that subjects in later sessions of the experiment might learn about the design before they participated the experiment, which could affect their behavior. To rule out this concern, we divide our experimental sessions into three waves, and compare the distribution of the senders' types across these waves. Wave 1-3 include the sessions conducted in the very first day of the experiment in 2021, later days in 2021, and all days in 2022, respectively. If senders in later waves learned about the surprise round, they would have additional strategic incentive to choose Nice, compared to senders in wave 1. Nevertheless, we find that the distributions of the senders' types do not vary significantly across the three waves ($p = 0.509$, Kruskal–Wallis tests of equality-of-populations).[9] Therefore, we rule out this concern.

3.2 Reciprocity level

Table 1 presents the average reciprocity level of the receiver by the sender's type and the receiver's realized outcome. Conditioned on the receiver receiving an outcome of 15, the

[9]The detailed distributions by waves and pairwise test results are provided in Table 4 in Appendix B.

reciprocity levels are always very low, no matter the sender is selfish (type-a) or somewhat nice (type-b). There is no significant difference between the two cases. Why receivers allocate almost nothing to the sender may be attributed to the large payoff inequality (whenever the receiver receives 15, the sender receives at least 28). Conditioned on the receiver receiving an outcome of 30, the average reciprocity level is 0.92 if the sender is "selfish" (type-a), it increases to 3.68 if the sender is "somewhat nice" (type-b), and it further increases to 4.26 if the sender is "very nice" (type-c). In summary, we find that, conditioned on receiving the same outcome, the reciprocity level almost always significantly increases in the sender's kindness level. This largely supports Hypothesis 1.

Table 1: Receiver's reciprocity level by outcome and the sender's type.

Outcome	Type-a	Type-b	Type-c	Diff test
15	0.40	0.33	–	$p = 0.923$ (a vs.b)
	(n=47)	(n=9)		
30	0.92	3.68	4.26	$p < 0.001$ (a vs. b)
	(n=26)	(n=19)	(n=58)	$p < 0.001$ (a vs. c)
				$p = 0.034$ (b vs. c)
Diff test	$p = 0.010$	$p < 0.001$		

Notes: Each cell shows the average amount allocated to the sender in the surprise round. The number of observations are in parentheses. Two-sided Mann-Whitney tests are performed to test the differences.

Result 1. Conditioned on the same outcome, a kinder sender type almost always induces a higher reciprocity level of the receiver. That is, reciprocity level $c30 > b30 > a30$, and $b15 = a15$.

Next, we compare the reciprocity level when senders make exactly the same three choices but the receivers receive a different outcome because of chance. First, when the sender is selfish (type-a), the receivers allocate significantly more if they receive 30 instead of 15. Second, when the sender is somewhat nice (type-b), the receivers also allocate

significantly more to the sender if they receive 30. In summary, we find that, conditioned on facing the same type of senders, the reciprocity level increases significantly in the receiver's payoff. This is in support of Hypothesis 2.

Result 2. Conditioned on the same sender type, a higher outcome of the receiver induces a higher reciprocity level of the receiver. That is, reciprocity level $a30 > a15$, and $b30 > b15$.

Finally, in our experiment, another factor may potentially affect the reciprocity level of the receiver. When a sender is a type-c, the final payoff will always be $(30, 30)$, no matter which game is chosen by the computer. Though the chosen game is a mere random choice by the computer, it may still affect the receiver's reciprocal behavior through the following behavioral channel. For example, suppose the receiver learns that the sender is a type-c and the first game ($x = 28$) is chosen for payment. However, in the first game, the final outcome would still be $(30, 30)$ even if the sender is a type-a or a type-b. This suggests that the sender does not need to be a type-c for the receiver to receive the desirable outcome $(30, 30)$. In contrast, if the third game ($x = 35$) is chosen, only a type-c sender can lead to the outcome $(30, 30)$. In other words, compared to when game 3 is chosen, the kindness level of the sender has a weaker causal relationship with the outcome $(30, 30)$ when game 1 is chosen. As a result, the receiver may feel less motivated to reciprocate the sender when the first game is chosen, compared to when the second game or the third game is chosen. Suppose this behavioral channel takes place, we would expect that, conditioned on having a type-c sender, the receiver may reciprocate more when the chosen game is game 3 ($x = 35$) than when it is game 2 ($x = 32$) than when it is game 1 ($x = 28$). Similarly, conditioned on having a type-b sender and receiving the same outcome $(30, 30)$ when game 1 or 2 is chosen, the receiver may reciprocate more when the chosen game is game 2 ($x = 32$) than when it is game 1 ($x = 28$).

Table 2 presents the average reciprocity levels of the receivers under different chosen games, conditioned on both the same sender type and the same outcome. In the case of $b30$, the reciprocity level does not differ significantly when the chosen game varies. In the case of $c30$, we do not find that the receivers reciprocate more under game 3 than

Table 2: Receiver's reciprocity level by the game chosen.

Type & Outcome	Game 1 ($x = 28$)	Game 2 ($x = 32$)	Game 3 ($x = 35$)	Diff test
$b30$	3.80	3.64	–	$p = 0.698$ (Game 1 vs. 2)
	(n=5)	(n=14)		
$c30$	5.00	3.47	4.23	$p = 0.005$ (Game 1 vs. 2)
	(n=17)	(n=15)	(n=26)	$p = 0.161$ (Game 1 vs. 3)
				$p = 0.141$ (Game 2 vs. 3)

Notes: Each cell shows the average amount allocated to the sender in the surprise round. The number of observations are in parentheses. Two-sided Mann-Whitney tests are performed to test the differences.

game 1 and 2, nor do we find that they reciprocate more under game 2 than game 1.[10] This result suggests that, upon learning the type of the sender and the final outcome, the receivers' reciprocal behaviors are not affected by the causal relationship between the sender type and the final outcome. This result further strengthens our result 1 on the effect of intention. That is, conditioned on the same outcome, the intention of the sender *per se* affects the reciprocity level, no matter it directly leads to the outcome or not. One real world example is that, when a colleague helps you to debug your code when she has an urgent deadline of her own task, you will view her intention as very kind (kinder than if she does not have an urgent deadline); and your view will not change even if the deadline of her task got postponed after she helped you.

Result 3. Conditioned on the same sender type and the same outcome, the causal relationship between the sender type and the outcome, which is reflected by the game chosen, does not affect the reciprocity level.

[10]The receivers actually reciprocate significantly more under game 1 than game 2, which cannot be explained by any reasonable behavioral mechanisms, although it may be attributed to a relatively small number of observations here ($n = 17$ and $n = 15$ under game 1 and 2, respectively).

To conclude, we find that a kinder sender induces a higher reciprocity level conditioned on the same outcome, a better outcome (for the receiver) induces a higher reciprocity level conditioned on the same kindness level of the sender, and the causal relationship between the sender type and the outcome does not affect the reciprocity level conditioned on the same sender type and outcome.

4 Theory

4.1 Competing models

4.1.1 Rabin (1993) and Dufwenberg and Kirchsteiger (2004)

Rabin [1993] proposes perhaps the first model of reciprocity in normal form games. It is developed in the framework of psychological game theory pioneered by Geanakoplos et al. [1989], which is further extended to extensive form games by Dufwenberg and Kirchsteiger [2004].

Higher order beliefs play a central role in psychological game theory. Although our design effectively eliminate higher order beliefs, we can still use the definition of kindness and reciprocity in Rabin [1993] and Dufwenberg and Kirchsteiger [2004].

The kindness function of the sender (player 1) to the receiver (player 2) is given by $k_{12}(S) = 15 - 22.5 = -7.5$, $k_{12}(N) = 30 - 22.5 = 7.5$, where S (N) denotes the Selfish (Nice) action and $22.5 = (15+30)/2$ is called the equitable payoff (what is fair) that is used to measure player 1's kindnesses toward player 2 corresponding to different choices.[11] One can observe that how much player 1 has to sacrifice is irrelevant to the kindness of player 1 defined here.

The kindness function of player 2 to player 1 is given by $k_{21}(t) = t - 5$, where $t \in [0, 10]$ denote the amount of money allocated by player 2 for player 1, and $5 = (0+10)/2$ is the equitable payoff that is used to measure player 2's kindnesses corresponding to different

[11] Note that, when making a choice between S and N in the first stage of the experimental game, the sender is not aware of the allocation game in the surprise round. Therefore the additional 10 does not enter the kindness function.

choices. Let Y_{21} measure player 2's sensitivity of reciprocity toward player 1. Assume $Y_{21} > 0$.

Player 2's utility function if player 1 chooses S is given by

$$U_2(t, S) = 15 + 10 - t + Y_{21} k_{21}(t) k_{12}(S) = 25 - t - 7.5 Y_{21}(t - 5). \tag{1}$$

Since it is a linear function in t, Player 2's optimal choice is $t = 0$.

Player 2's utility function if player 1 chooses N is given by

$$U_2(t, N) = 30 + 10 - t + Y_{21} k_{21}(t) k_{12}(N) = 40 - t + 7.5 Y_{21}(t - 5). \tag{2}$$

Player 2's optimal choice is $t = 10$ if Y_{21} is sufficiently large, and is $t = 0$ otherwise.

The models of Rabin [1993] and Dufwenberg and Kirchsteiger [2004] predict that player 2's reciprocity toward player 1 depends solely on player 2's payoff. In other words, the difference between S and N in terms of kindness is identical across the three types of player 1: a, b and c, provided that player 2 receives the same payoff.

4.1.2 Charness and Rabin (2002) and Falk and Fischbacher (2006)

Whereas Rabin [1993] and Dufwenberg and Kirchsteiger [2004] concentrate on modeling the general principles of reciprocity, Charness and Rabin [2002] and Falk and Fischbacher [2006] try to combine distributional preferences and psychological game theoretical based reciprocity into unifying models. Importantly, both papers take into account the role of sacrifice. Falk and Fischbacher [2006] assume that a player does not resent harmful behavior by the other player if it seems to come only from the other player's unwillingness to come out behind rather than her selfishness when ahead. To make this point clearer, consider the example in their paper. In a dictator game, suppose the dictator can choose from two possibilities of outcome combinations (own, other): (8, 2) and (2, 8). The receiver may not resent (8, 2) too much because it is not reasonable to demand the dictator to be fair with the receiver since it implies that the dictator would put herself in a very disadvantageous position. Now, suppose a third option (5, 5) is available to the dictator, then (8, 2) will be perceived to be quite unkind by the receiver because (5, 5) is a more friendly offer than (8, 2) and it does not put the dictator in a disadvantageous position.

In sum, how unkind an action is perceived depends on how much the dictator has to sacrifice in order to make the more friendly offer. Alternatively, Charness and Rabin [2002] hypothesize that an action is deemed unkind if the decision maker chooses it by pursuing self-interest at the expense of social welfare preferences.

Both Charness and Rabin [2002] and Falk and Fischbacher [2006] focus on defining how sacrifice affects the degree of unkindness of an unkind action. For example, in our game, according to Falk and Fischbacher [2006], S given $x = 32$ should be perceived as more unkind than S given $x = 35$ because player 1 has to sacrifice more to make the more friendly offer (N) given $x = 35$.[12] Nevertheless, neither has a clear definition on how sacrifice is related to the degree of kindness of a kind action.[13]

4.1.3 Levine (1998)

Levine [1998] considers a parametric model of reciprocity that is not based on psychological game theory. Suppose there are two players, player 1 and 2. m is player 2's payoff and y is the player 1's payoff. Player 2's utility is given by:

$$U_2 = m + \frac{\alpha_2 + \lambda\alpha_1}{1 + \lambda} y. \tag{3}$$

Here α_2 is player 2's altruistic preference parameter, α_1 is player 2's estimation of player 1's altruistic preference parameter. When $\lambda = 0$, it is the purely altruistic model. When $\lambda > 0$, reciprocity kicks into play: player 2 is willing to be more altruistic toward player 1 if player 1 is more altruistic toward player 2.

Note that this model involves imperfect information about α_1, so player 2 needs to form belief about it and the game becomes a signaling game. Player 1 can strategically reveal information about his altruism to player 2. Our experimental design effective eliminates the strategic concern in this model, so the choices made by player 1 perfectly reveal his altruism to player 2.

[12] Whether S given $x = 32$ is more or less unkind than S given $x = 35$ depends on the definition of social welfare preferences in Charness and Rabin [2002]'s model.

[13] Note that according to Falk and Fischbacher [2006]'s model, player 1's kindness toward player 2 by choosing N is zero because player 2's payoff minus player 1's payoff (30-30) equals 0 in the first stage of the experimental game!

This model can potentially support Result 1 (Hypothesis 1) that $c30 > b30 > a30$, $b15 > a15$ because it allows for interpersonal comparison of altruism: type-c is more altruistic than type-b, which in turn is more altruistic than type-a.

However, this model fails to support our Result 2 (Hypothesis 2) that $b30 > b15$ and $a30 > a15$ because player 2's absolute payoff does not matter in this model due to its linearity. Whether a non-linear variation of the model can support our Result 2 (Hypthesis 2) is unclear. Even if we can find one, it may be difficult to justify why non-linearity is needed to produce the desired results. Hence, a non-parametric model may be preferable.

4.1.4 Cox, Friedman and Sadiraj (2008)

Cox et al. [2008] proposes an axiomatic model of reciprocity. To fix idea, consider the simplest utility function involving altruism for two players (me and the other):

$$U(m, y) = m + \alpha y, \tag{4}$$

where m is my own payoff, y is the other's payoff, α is my altruism parameter (here α is not necessarily a fixed preference parameter, but a function of factors such as the other's intention or kindness level).

Willingness to pay, $WTP(m.y)$, the amount of own payoff I am willing to give up in order to increase the other player's payoff by a unit, is exactly α. Note that given general utility function form involving tradeoff between own payoff and the other's payoff, one can always measure WTP. Consider preference orderings in terms of the trade-off between my own payoff and the other's payoff that are smooth and convex in R^2_+ and strictly increasing in my own payoff:

Definition 1. *Preference ordering A is said to be more altruistic than* ***(MAT)*** *B if* $WTP_A(m.y) \geq WTP_B(m.y)$.

Define an opportunity set F as a subset of R^2. Each element of F is a vector on my own payoff m and the other's payoff y.

Definition 2. *Opportunity set G is said to be more generous than* ***(MGT)*** *another opportunity set F if (1)* $m^*_G - m^*_F \geq 0$ *and (2)* $m^*_G - m^*_F \geq y^*_G - y^*_F$.

m^*_G (respectively, y^*_G) is the highest payoff I (respectively, the other player) can get from the opportunity set G. Hence, criterion (1) in Definition 2 says that G is more generous than F if I can get a higher payoff from G and criterion (2) in Definition 2 says that this is true as long as the other player does not increase her own potential payoff more than mine.

Consider a two-player extensive form game with complete information in which player 1 (the first mover) chooses an opportunity set C which belongs to a set of possible opportunity sets $\mathcal{C}$ that he can choose from, and the second mover chooses the payoffs $(m, y) \in C$. Initially, player 2 knows $\mathcal{C}$. Prior to her choice of payoffs, player 2 learns the actual opportunity set C, and acquire preferences A_C.

Axiom R. Suppose player 1 chooses the actual opportunity set for player 2. If $F, G \in \mathcal{C}$ and $G\ MGT\ F$, then $A_G\ MAT\ A_F$.

This axiom is very intuitive. It says that if player 1 is more generous to player 2, than player 2 will be more altruistic toward player 1, Note that this axiom is also implicitly considered in Cox et al. [2007]'s parametric model of reciprocity.

Now, let us apply this axiomatic model to the sender-receiver game shown in Figure 3 with x taking values from $\{28, 32, 35\}$, which essentially corresponds to the three games considered in our experiment without the surprise round element.

Let $\tilde{S}(\tilde{N})$ be the opportunity set induced by action $S(N)$ by player 1. $\tilde{N}$ indeed $MGT\ \tilde{S}$ because 1) $m^*_{\tilde{N}} - m^*_{\tilde{S}} = 30 + 10 - (15 + 10) = 15$ and 2) $m^*_{\tilde{N}} - m^*_{\tilde{S}} = 15 \geq y^*_{\tilde{N}} - y^*_{\tilde{S}} = (30 + 10) - (x + 10) = 30 - x$, for $x = 28, 32, 35$.

Hence, according to Axiom R, $A_{\tilde{N}}\ MAT\ A_{\tilde{S}}$. That is, player 2 is more altruistic toward player 1 when player 1 chooses N. However, there is no additional information on whether player 2 would feel differently toward the Nice action chosen by player 1 given the three different outside options $28, 32, 35$.

The limitation of this model is on the definition of MGT (more generous than). Although criterion (2) in Definition 2 involves the consideration of player 1's payoff, it is silent about sacrifice because when player 1 makes a sacrifice, we have $y^*_{\tilde{N}} - y^*_{\tilde{S}} < 0$, and criterion (1) directly implies criterion (2)!

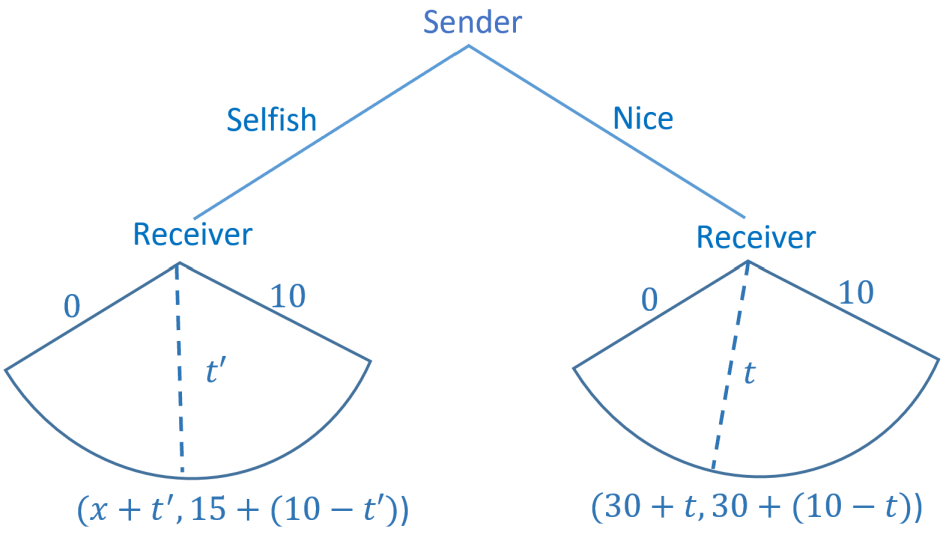

Figure 3: A Sender-Receiver Game

4.2 A simple model of reciprocity incorporating sacrifice

In this section, we extend Cox et al. [2008]'s model to take in account the potential effect of sacrifice. More specifically, Cox et al. [2008]'s approach lacks a measure of the intensity of generosity, so we add the following additional definition:

Definition 3. *Consider four opportunity sets E, F, G, H and suppose H MGT F, G MGT E. We say H MGT F more than G MGT E if (1) $m_H^* - m_F^* \geq m_G^* - m_E^*$ and (2) $y_G^* - y_E^* \geq y_H^* - y_F^*$.*

Definition 3 provides an intuitive measure of the intensity of generosity that can account for sacrifice. Criterion (1) in Definition 3 says that I get a weakly higher payoff gain from opportunity set H over F, than from G over E. Criterion (2) says that the other player sacrifices more (gains less) by choosing H over F, than by choosing G over E. In the game shown in Figure 3, let $\tilde{N}$ denote the opportunity set followed by player 1's choice of Nice; $\tilde{S}_x$ denote the opportunity set followed by player 1's choice of Selfish when his outside option is x. Then we have $\tilde{N}$ MGT $\tilde{S}_{35}$ more than $\tilde{N}$ MGT $\tilde{S}_{32}$ more than $\tilde{N}$ MGT $\tilde{S}_{28}$. In the context of our experiment, we can obtain interpersonal comparisons of generosity for the subjects in the spirit of Levine [1998] but with a solid axiomatic foundation: a type-c player 1 (represented by a vector of opportunity sets $(\tilde{N}, \tilde{N}, \tilde{N})$) is more generous than a type-b player 1 (represented by $(\tilde{N}, \tilde{N}, \tilde{S}_{35})$), who in turn is more

generous than a type-a player 1 (represented by $(\tilde{N}, \tilde{S}_{32}, \tilde{S}_{35})$).

In Appendix C1, we show that Definition 3 provides a well-defined partial ordering on an appropriately defined set of MGT relations.

We add the following axiom in addition to Axiom R to capture how an individual reacts to different intensities of generosity:

Axiom R'. Suppose $H, F \in \mathcal{C}_1 = \{H, F\}$ and $G, E \in \mathcal{C}_2 = \{G, E\}$. If H MGT F more than G MGT E, then $A_H \; MAT \; A_G$.

In the game shown in Figure 3, Axiom R' states that player 2 is more altruistic toward player 1 when he needs to make a bigger sacrifice to choose the Nice action.

Now let us consider the choice made by player 2, who is trying to choose her most preferred point in her opportunity set as assumed in the neoclassical theory. For notation convenience, let $\tilde{N}_x$ denote the opportunity set followed by player 1's choice of Nice when his outside option is x. Mathematically, $\tilde{N}_{35} = \tilde{N}_{32} = \tilde{N}_{28} = \tilde{N}$. Let $(m_{\tilde{N}_x}, y_{\tilde{N}_x})$ denote a $A_{\tilde{N}_x}$-chosen point in $\tilde{N}$. Suppose the induced preference orderings $A_{\tilde{N}_x}$ are strictly convex, according to Proposition 1 of Cox et al. [2008], $(m_{\tilde{N}_x}, y_{\tilde{N}_x})$ is unique. Then according to Part 1 of Proposition 2 of Cox et al. [2008], since by Axiom R', $A_{\tilde{N}_{35}}$ MAT $A_{\tilde{N}_{32}}$ MAT $A_{\tilde{N}_{28}}$, we have $y_{\tilde{N}_{35}} > y_{\tilde{N}_{32}} > y_{\tilde{N}_{28}}$. In other words, player 2 would reward more to player 1 when he needs to make a bigger sacrifice to choose the Nice action.

In the context of our experiment, Axiom R' would intuitively imply that player 2 rewards more to a type-c player 1 than a type-b player 1 than a type-a player 1. Hence, Axiom R' matches our Result 1 and Hypothesis 1. Note that with Axiom R, the model is still compatible with Result 2 and Hypothesis 2. However, Axiom R' does not allow us to directly compare $b15$ and $a30$ for example, because we do not have clear empirical guidance on the trade-off between intention and outcome.

Note that Definition 2, Definition 3, Axiom R and Axiom R' together are consistent with some choice function representation. Let us consider the CES function specified in

Cox et al. [2007]:

$$u_J(m,y) = \begin{cases} (m^\alpha + \theta_J y^\alpha)/\alpha & \text{if } \alpha \in (-\infty, 1)\backslash\{0\}, \\ m y^{\theta_J} & \text{if } \alpha = 0, \end{cases}$$

where $\alpha \in (-\infty, 1)$ is the convexity parameter, and θ_J is the willingness to pay own for other's payoff when $m = y$ given that the opportunity set faced by me is J. It is straightforward to verify that if $0 < \theta_K \leq \theta_J$, then A_J MAT A_K.

Suppose J belongs to a set of opportunity sets $\mathcal{C}$. Let $\theta_J = \theta(g(J))$, where $\theta(\cdot)$ is an increasing function of the generosity of J, which is measured by $g(J) = (m_J^* - m_N^*) - (y_J^* - y_N^*)$, where N is a set chosen from $\mathcal{C}$ to serve as the baseline. If J MGT K for some $K \in \mathcal{C}$, by Definition 2, we have $g(J) - g(K) = (m_J^* - m_K^*) - (y_J^* - y_K^*) \geq 0$. Hence, $\theta_J \geq \theta_K$, which is consistent with Axiom R.

Now, consider two binary sets of opportunity sets, $\mathcal{C}_1 = \{H, F\}$ and $\mathcal{C}_2 = \{G, E\}$. Suppose H MGT F more than G MGT E, let F and E be the baseline sets for $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.[14] By Definition 3, we have $g(H) = (m_H^* - m_F^*) - (y_H^* - y_F^*) \geq g(G) = (m_G^* - m_E^*) - (y_G^* - y_E^*)$. Hence, $\theta_H \geq \theta_G$, which is consistent with Axiom R".

In Appendix C2, we show that Definition 3 provides a partial ordering of feasible sets in some existing experimental games and Axiom R' can be applied to some existing experimental data.

Axiom R' may seem restrictive as we require $\mathcal{C}_1$ and $\mathcal{C}_2$ to be binary. We do this because MGT defined in Definition 2 only gives a partial ordering for the opportunity sets. Note that Cox et al. [2008] provide a lite version of MGT which only requires criterion (1) in Definition 2 and this MGT-Lite gives a complete ordering for the oppurtunity sets. We can thus replace MGT with MGT-Lite in Definition 3 and generalize Axiom R' as follows:

Axiom R". Consider two finite sets of opportunity sets $\mathcal{C}_1, \mathcal{C}_2$. Let F (E) be the least generous opportunity set in $\mathcal{C}_1$ $(\mathcal{C}_2)$ according to MGT-Lite. For $H \in \mathcal{C}_1$, $G \in \mathcal{C}_2$, if H MGT-Lite

[14] Cox et al. [2007] do not specify what the choice of the baseline should be. Instead, they treat it as an empirical question. In the specific context we consider here, however, since both sets of opportunity sets are binary, we think it is reasonable to treat F and E as the baseline sets as each of them is the least ranked in terms of MGT in its respective set of opportunity sets.

F more than G MGT-Lite E, then A_H MAT A_G.

To make two opportunity sets that belong to two different non-binary sets of opportunity sets comparable, we first need to determine the baseline in each set of opportunity sets. We argue that a natural baseline that I would consider for a set of opportunity sets is its least generous opportunity set (which always exists by using MGT-Lite to rank all the sets in terms of generosity) because it gives me the smallest monetary payoff possible. Axiom R" states that when comparing two opportunity sets that belong to two different sets of opportunity sets, if one gives me more and requires a larger sacrifice from the other person comparing to the baseline in its respective set of opportunity sets than the other opportunity does, then I should feel more altruistic toward the other person given the former than the latter.

Although MGT-Lite provides a well-defined complete ordering, MGT-Lite more than MGT-Lite we specify in Axiom R" is not complete on the set collecting all MGT-Lite relations. Consider two finite sets of opportunity sets:

$$\mathcal{C}_1 = \{\{(3,0),(0,0)\},\{(1.5,1.5),(0,0)\},\{(0,0),(0,0)\}\},$$
$$\mathcal{C}_2 == \{\{(2,0),(0,0)\},\{(1,1),(0,0)\},\{(0,0),(0,0)\}\}.$$

$F = \{(0,0),(0,0)\}$ (respectively, $E = \{(0,0),(0,0)\}$) are the least generous opportunity set in $\mathcal{C}_1$ (respectively, $\mathcal{C}_2$) according to MGT-Lite. Now consider $H = \{(1.5,1.5),(0,0)\} \in \mathcal{C}_1$, $G = \{(1,1),(0,0)\} \in \mathcal{C}_2$, which satisfy H MGT-Lite F, G MGT-Lite E, and $m^*_H - m^*_F = 1.5 > m^*_G - m^*_E = 1$. However, $y^*_G - y^*_E = 1 < y^*_H - y^*_F = 1.5$. In this case, H MGT-Lite F and G MGT-Lite E are not comparable. Nevertheless, in some special examples, such as a constant sum game, MGT-Lite more than MGT-Lite can be complete.

5 Conclusion

This paper experimentally examines the role of sacrifice in reciprocity. We employ a novel design, which allows us to obtain a clean measurement of the sender's kindness in terms of his sacrifice in a combination of three sender-receiver games and we show that the receiver

has a stronger tendency to reciprocate the sender if the sender is willing to sacrifice more. We also show that conditioned on the same level of kindness of the sender, the receiver tends to reciprocate more when she gets a better outcome, which matches the conventional wisdom shared by most of the existing theories. Finally, we propose a simple model of reciprocity to accommodate sacrifice.

The theory of reciprocity has been extensively used to understand various applied economic problems including wage undercutting [Dufwenberg and Kirchsteiger, 2000], voting [Hahn, 2009], climate negotiations [Nyborg, 2018], public good investment [Dufwenberg and Patel, 2017, Jang et al., 2018, Kozlovskaya and Nicoló, 2019], randomized policy experiments [Aldashev et al., 2017], performace based contracts [Livio and De Chiara, 2019], trade disputes [Conconi et al., 2017], mechanism design [Bierbrauer and Netzer, 2016, Bierbrauer et al., 2017], trust-based lending relationship [Hyndman et al., 2021], conflict of interests in third party reviews [Ham et al., 2021]. insolvency in banking [Dufwenberg and Rietzke, 2020], among many others. We hope that the role of sacrifice can be further explored in applications.

References

G. A. Akerlof. Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, 97: 543–569, 1982.

G. A. Akerlof and J. L. Yellen. Fairness and unemployment. *American Economic Review*, 78: 44–49, 1988.

G. A. Akerlof and J. L. Yellen. The fair-wage effort hypothesis and unemployment. *Quarterly Journal of Economics*, 105:255–284, 1990.

G. Aldashev, G. Kirchsteiger, and A. Sebald. Assignment procedure biases in randomized policy experiments. *Economic Journal*, 127:873–895, 2017.

P. Battigalli and M. Dufwenberg. Belief-dependent motivations and psychological game theory. *Journal of Economic Literature*, Forthcoming, 2021.

F. Bierbrauer and N. Netzer. echanism design and intentions. *Journal of Economic Theory*, 163:557–603, 2016.

F. Bierbrauer, A. Ockenfels, A. Pollak, and D. Rückert. Robust mechanism design and social preferences. *Journal of Public Economics*, 149:59–80, 2017.

G. E. Bolton and R. Zwick. Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior*, 10:95–121, 1995.

J. Brandts, E. Fatas, E. Haruvy, and F. Lagos. The impact of relative position and returns on sacrifice and reciprocity: An experimental study using individual decisions. *Social Choice and Welfare*, 45:489–511, 2014.

B. Çelen, A. Schotter, and M. Blanc. On blame and reciprocity: Theory and experiments. *Journal of Economic Theory*, 169:62–92, 2017.

G. Charness and P. Kuhn. Lab labor: What can labor economists learn from the lab? In *Handbook of Labor Economics*, volume 4A, chapter 3, pages 229–330. Elsevier, 2011.

G. Charness and M. Rabin. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117:817–869, 2002.

I. Cho, H.-J. Song, H. Kim, and S. Sul. Older adults consider others' intentions less but allocentric outcomes more than young adults during an ultimatum game. *Psychology and Aging*, 35(7):974–980, 2020.

P. Conconi, D. R. DeRemer, G. Kirchsteiger, L. Trimarchi, and M. Zanardi. Suspiciously timed trade disputes. *Journal of International Economics*, 105:57–75, 2017.

J. C. Cox, D. Friedman, and S. Gjerstad. A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59:17–45, 2007.

J. C. Cox, D. Friedman, and V. Sadiraj. Revealed altruism. *Econometrica*, 76:31–69, 2008.

M. Dufwenberg and G. Kirchsteiger. Reciprocity and wage undercutting. *European Economic Review*, 44:1069–1078, 2000.

M. Dufwenberg and G. Kirchsteiger. A theory of sequential reciprocity. *Games and Economic Behavior*, 47:268–298, 2004.

M. Dufwenberg and A. Patel. Reciprocity networks and the participation problem. *Games and Economic Behavior*, 101:260–272, 2017.

M. Dufwenberg and D. Rietzke. Banking on reciprocity: Deposit insurance and insolvency. Mimeo, 2020.

M. Dufwenberg, A. Smith, and M. Van Essen. Hold-up: With a vengeance. *Economic Inquiry*, 51:896–908, 2013.

A. Falk and U. Fischbacher. A theory of reciprocity. *Games and Economic Behavior*, 54: 293–315, 2006.

A. Falk, E. Fehr, and U. Fischbacher. On the nature of fair behavior. *Economic Inquiry*, 41: 20–26, 2003.

E. Fehr, G. Kirchsteiger, and A. Riedl. Does fairness prevent market clearing?: An experimental investigation. *Quarterly Journal of Economics*, 108:437–459, 1993.

U. Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178, 2007.

J. Gale, K. G. Binmore, and L. Samuelson. Learning to be imperfect: The ultimatum game. *Games and Economic Behavior*, 8:56–90, 1995.

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

W. Güth and M. G. Kocher. More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior and Organization*, 108:396–409, 2014.

W. Güth, R. Schmittberger, and B. Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3:367–388, 1982.

V. Hahn. Reciprocity and voting. *Games and Economic Behavior*, 67:467–480, 2009.

S. H. Ham, I. Koch, N. Lim, and J. Wu. Conflict of interest in third-party reviews: An experimental study. *Management Science*, Forthcoming, 2021.

K. Hyndman, J. Wu, and S. C. Xiao. Trust and lending: An experimental study. Mimeo, 2021.

D. Jang, A. Patel, and M. Dufwenberg. Agreements with reciprocity: Co-financing and mous. *Games and Economic Behavior*, 111:85–99, 2018.

L. Jiang and J. Wu. Belief-updating rule and sequential reciprocity. *Games and Economic Behavior*, 113:770–780, 2019.

A. A. Johnsen and O. Kvaløy. Does strategic kindness crowd out prosocial behavior? *Journal of Economic Behavior and Organization*, 132:1–11, 2016.

M. Kozlovskaya and A. Nicoló. Public good provision mechanisms and reciprocity. *Journal of Economic Behavior and Organization*, 167:235–244, 2019.

D. Levine. Modeling altruism and spitefulness in experiment. *Review of Economic Dynamics*, 1:593–622, 1998.

L. Livio and A. De Chiara. Friends or foes? optimal incentives for reciprocal agents. *Journal of Economic Behavior and Organization*, 167:245–278, 2019.

K. A. McCabe, M. L. Rigdon, and V. L. Smith. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, 52(2):267–275, 2003.

K. Nyborg. Reciprocal climate negotiators. *Journal of Environmental Economics and Management*, 92:707–725, 2018.

Y. A. Orhun. Perceived motives and reciprocity. *Games and Economic Behavior*, 109: 436–451, 2018.

M. Rabin. Incorporating fairness into game theory and economics. *American Economic Review*, 83:1281–1302, 1993.

A. Sebald. Attribution and reciprocity. *Games and Economic Behavior*, 68:339–352, 2010.

J. Sobel. Interdependent preferences and reciprocity. *Journal of Economic Literature*, 43: 392–436, 2005.

J. Sohn and W. Wu. Reciprocity with uncertainty about others. Mimeo, 2021.

L. Stanca, L. Bruni, and L. Corazzini. Testing theories of reciprocity: do motivations matter? *Journal of Economic Behavior and Organization*, 71:233–245, 2009.

M. Sutter. Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology*, 28:69–78, 2007.

M. Wittig, K. Jensen, and M. Tomasello. Five-year-olds understand fair as equal in a mini-ultimatum game. *Journal of Experimental Child Psychology*, 116:324–337, 2013.

Appendices

Appendix A: Experimental instructions

In this appendix, we provide the experimental instructions and the experimental screenshots that are translated from the original Chinese version. Figure 4 presents a screenshot of the experiment.

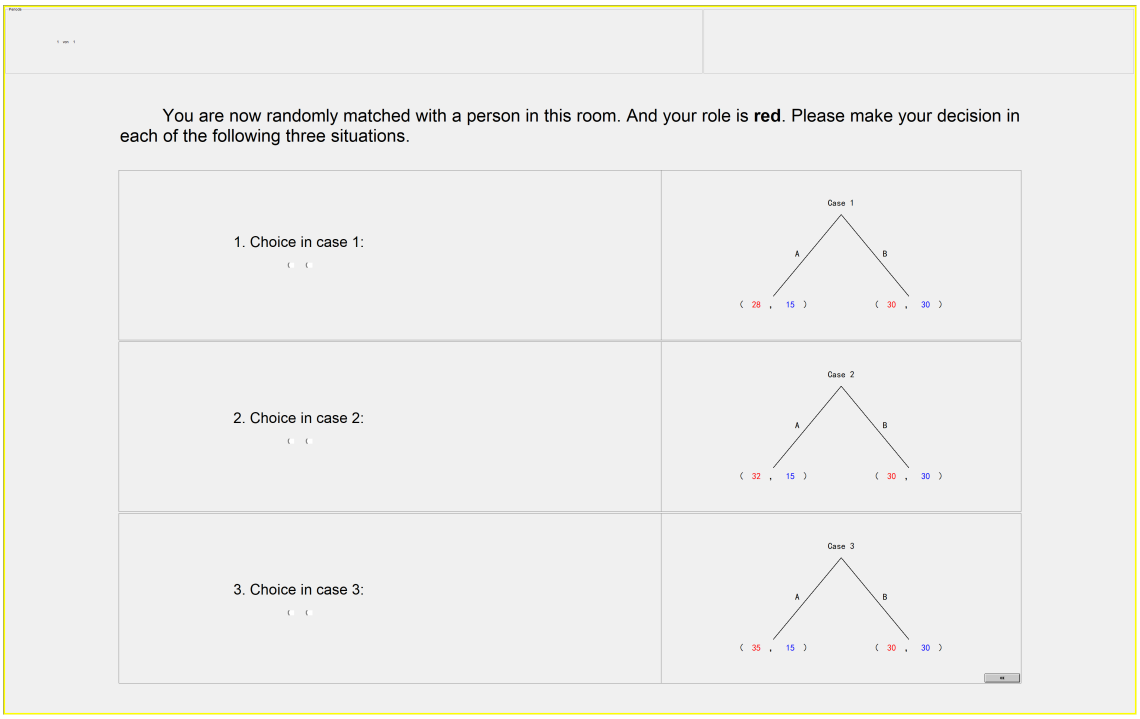

Figure 4: A (translated) screenshot of the experiment

Instructions (for all players)

Welcome to this experiment on decision-making. Please read the following instructions carefully. During the experiment, do not communicate with other participants in any means. If you have any question at any time, please raise your hand, and an experimenter will come and assist you privately. This experiment will last about half an hour. You are going to take part in an experiment in this room together with other participants. Each participant seat behind a private computer, and no one can ever know the identity of another. It is an anonymous experiment. Experimenters and other participants cannot

link your name to your desk number, and thus will not know the identity of you or of other participants who made the specific decisions. Your earnings are denoted in "RMB (Yuan)" throughout the experiment. Your earnings may depend on your own choices and the choices of other participants. In addition, you receive 15 Yuan as show-up fee. This show-up fee is added to your earnings from the experiment. Your total earnings will be paid to you in cash privately.

In this experiment, you will be matched with another subject in the room. You and your matched partner will be randomly assigned a color: red or blue. If you are red, your partner must be blue. If you are blue, your partner must be red. In each pair, the red one can make some choices that affect both subjects' earnings, while the blue one does not have a choice. If you are red, you will need to make in total three decisions, one in each of the following three cases.

Case 1: The red can choose between two choices: A and B. If the red chooses A, the red gets 28 Yuan and the partner blue gets 15 Yuan. If the red chooses B, both the red and the blue get 30 Yuan. The choices are represented in the figure below. In the figure, the red number is the payoff of the red, and the blue number is the payoff of the blue.

Case 2: The red can choose between two choices: A and B. If the red chooses A, the red gets 32 Yuan and the partner blue gets 15 Yuan. If the red chooses B, both the red and the blue get 30 Yuan.

Case 3: The red can choose between two choices: A and B. If the red chooses A, the red gets 35 Yuan and the partner blue gets 15 Yuan. If the red chooses B, both the red and the blue get 30 Yuan.

After the red finishes making the three decisions, the computer will randomly draw ONE case out of the three (the probability of drawing any case is identical, that is, one third for each case). And the red player's previous decision in this randomly drawn case will uniquely determine the payoff of both the red and the blue in that pair. Then, the game ends, and the payoffs are realized.

For example, suppose a red chooses A in case 1, B in case 2, and A in case 3. After he or she makes such choices, suppose the computer randomly draws case 3. Then the final payoff is determined by the red player's choice A in case 3, which yields 35 Yuan for red,

and 15 Yuan for blue.

After the computer makes a random draw, both the red and the blue will be informed of which of the three cases is drawn and their realized payoffs in this case. The blue one will also be informed of each of the decisions made by the red in ALL the three cases.

Surprise Round (only for blue players and after the red players made their choices)

Now, as a blue player, you enter a surprise stage. Note that, all the information in this stage is never mentioned in the instructions of the experiment. Also, your partnered red player is not aware of this stage even until now.

We now give your 10 Yuan in total to allocate between yourself and your partner red. You can allocate these 10 Yuan only in integer numbers. Please decide your allocation below. After you make this allocation, the red player will learn her extra earnings (as a surprise!), and then the experiment will end definitely.

Appendix B: Additional tables.

Table 3: Demographic information by roles.

	Female, %	Age	Grade
Senders	73.8%	20.7	2.03
	(3.45)	(0.138)	(0.119)
Receivers	72.6%	20.9	2.29
	(3.49)	(0.137)	(0.128)
p-value	0.804	0.335	0.154

Notes: Standard errors in parentheses. p-values refer to Kruskal–Wallis tests of equality-of-populations between the roles of the subjects.

Table 4: Type distribution of the senders in three waves of the experiment.

Wave	Time	No. of sessions	No. of subjects	Type-a'	Type-a	Type-b	Type-c
1	First day in 2021	5	112	1.8%	46.4%	25.0%	26.8%
2	Later days in 2021	6	98	6.1%	44.9%	12.2%	36.7%
3	All days in 2022	4	118	1.7%	42.4%	13.6%	42.4%

Notes: Type-a' refers to the anti-social type, type-a refers to the selfish type, type-b is somewhat nice, and type-c is very nice. p-values of Kruskal–Wallis tests of equality-of-populations between different waves are 0.926 (wave 1 vs. 2), 0.271 (wave 1 vs. 3), and 0.385 (wave 2 vs. 3), respectively.

Appendix C1: Proof that MGT-MT-MGT is a Partial Ordering

Let S_{MGT} be the set collecting all MGT relations defined on the power set of $\mathbb{R}^+_2$. For example, suppose $F, G \subseteq \mathbb{R}^+_2$ and F MGT G, then $F-MGT-G$ is an element of S_{MGT}. Furthermore, we refine S_{MGT} by imposing the following restriction: $H-MGT-F$ is considered as identical to $G-MGT-E$ ($H-MGT-F$=$G-MGT-E$) if $m^*_H - m^*_F = m^*_G - m^*_E$ and $y^*_G - y^*_E = y^*_H - y^*_F$. The restriction effectively reduces the number of elements in S_{MGT} by treating MGT relations that describe the same level of relative generosity as the same. Let S^R_{MGT} denote the refined S_{MGT} .

Next, we prove that MGT-MT-MGT is a well-defined partial ordering on S^R_{MGT}.

- First, it satisfies reflexivity. Suppose H MGT F, then H MGT F more than H MGT F because $m^*_H - m^*_F \geq m^*_H - m^*_F$ and $y^*_H - y^*_F \geq y^*_H - y^*_F$.

- Second, it satisfies antisymmetry. Suppose H MGT F more than G MGT E, and G MGT E more than H MGT F. That is, $m^*_H - m^*_F \geq m^*_G - m^*_E$ and $y^*_G - y^*_E \geq y^*_H - y^*_F$; and $m^*_G - m^*_E \geq m^*_H - m^*_F$ and $y^*_H - y^*_F \geq y^*_G - y^*_E$. These inequalities imply that $m^*_H - m^*_F = m^*_G - m^*_E$ and $y^*_G - y^*_E = y^*_H - y^*_F$. Hence, by definition $H-MGT-F$=$G-MGT-E$ on S^R_{MGT}.[15]

- Third, it satisfies transitivity. Suppose H MGT F more than G MGT E, and G MGT E

[15]Note that antisymmetry is not satisfied if we consider S_{MGT}.

more than K MGT L, then $m^*_H - m^*_F \geq m^*_G - m^*_E \geq m^*_K - m^*_L$ and $y^*_K - y^*_L \geq y^*_G - y^*_E \geq y^*_H - y^*_F$, implying that H MGT F more than K MGT L.

Appendix C2: Diagnosing Definition 3 and Axiom R' Using Existing Experiments

Example 1: Ultimatum Mini-Game [Bolton and Zwick, 1995, Gale et al., 1995, Falk et al., 2003]

As in Cox et al. [2007], we use the data reported in Falk et al. [2003], who consider the four games in Figure 5. The authors employ the strategy method for the second player. Hence, 45 pairs of subjects generate 45 observations from each decision node in each game tree. The observed frequencies of all actions (in the parentheses under actions) are listed in Figure 5 as well.

Let a^{G_x} denote the opportunity set induced by action $a \in \{Take, Share\}$ in Game $x \in \{1, 2, 3, 4\}$. Observe that

- $m^*_{Take^{G_1}} - m^*_{Share^{G_1}} = 2$ and $y^*_{Take^{G_1}} - y^*_{Share^{G_1}} = -2$;
- $m^*_{Take^{G_2}} - m^*_{Share^{G_2}} = 0$ and $y^*_{Take^{G_2}} - y^*_{Share^{G_2}} = -0$;
- $m^*_{Take^{G_3}} - m^*_{Share^{G_3}} = -3$ and $y^*_{Take^{G_3}} - y^*_{Share^{G_3}} = 3$;
- $m^*_{Take^{G_4}} - m^*_{Share^{G_4}} = -6$ and $y^*_{Take^{G_4}} - y^*_{Share^{G_4}} = 6$;

According to Definition 3, we have $Share^{G_4}$ MGT $Take^{G_4}$ more than $Share^{G_3}$ MGT $Take^{G_3}$ more than $Take^{G_1}$ MGT $Share^{G_1}$ more than $Share^{G_2}$ MGT $Take^{G_2}$.

Hence, by Axiom R', we should have $A_{Share^{G_4}}$ MAT $A_{Share^{G_3}}$ MAT $A_{Take^{G_1}}$ MAT $A_{Share^{G_2}}$. The data is largely consistent with Axiom R' as 44 subjects chose Accept in G4, 45 subjects chose Accept in G3, 41 subjects chose Tolerate in G1, and 37 subjects chose Accept in G2.

In several subsequent experiments, similar patterns are found. Sutter [2007] runs the ultimatum mini-game specified in Figure 5 across three age groups (children, teens, university students) and he finds that on average 82% subjects chose Accept in G4; 99%

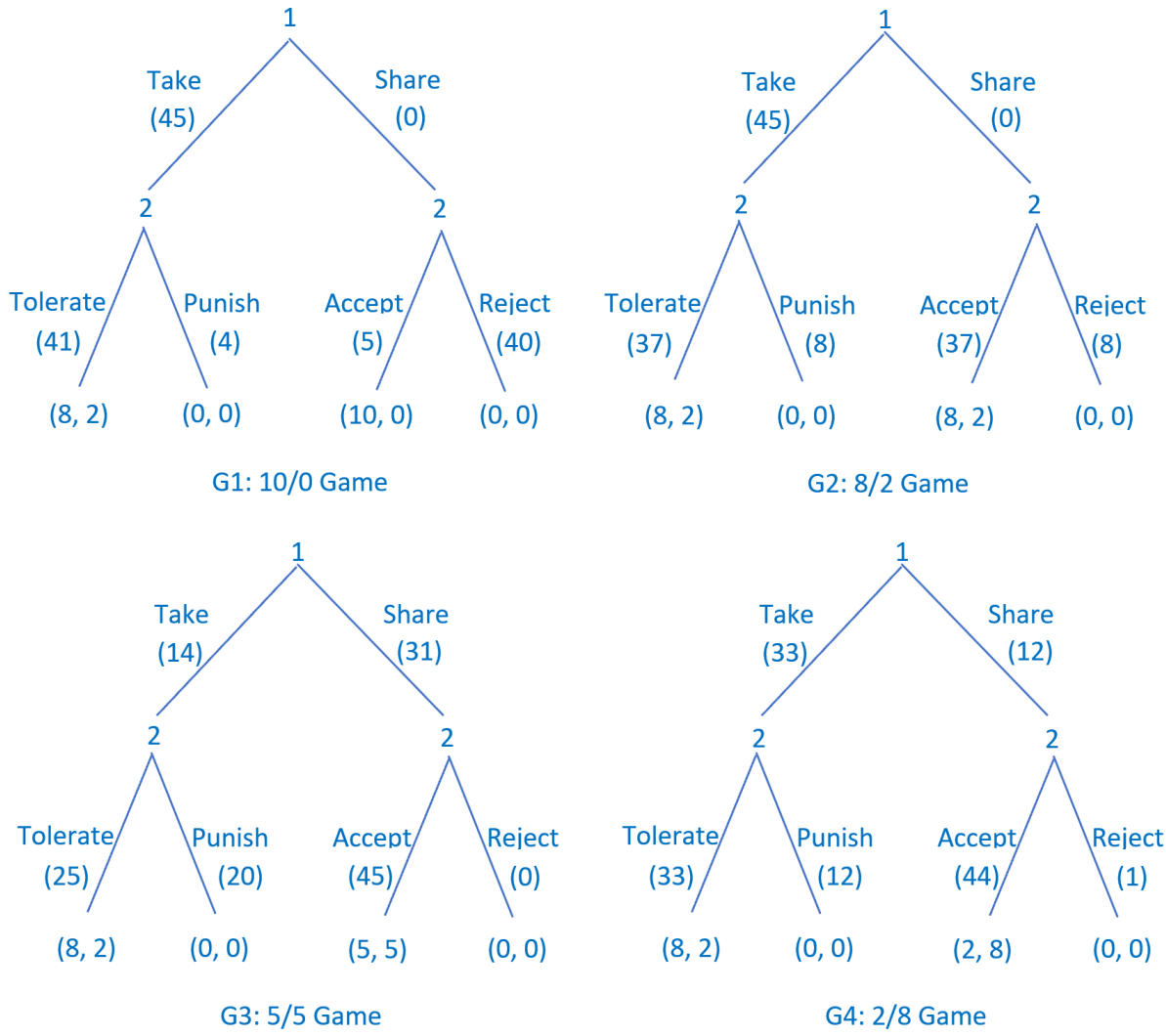

Figure 5: Ultimatum Mini-Games

subjects chose Accept in G3; 66% subjects chose Tolerate in G1; 63% subjects chose Accept in G2.

Wittig et al. [2013] test four ultimatum-mini games similar to those in Figure 5 in a group of children. In the four games they consider, the common payoff $(8,2)$ across games is replaced by $(3,1)$; $(10,0)$ in G1 is replaced by $(4,0)$; $(5,5)$ in G3 is replaced by $(2,2)$; and $(2,8)$ in G4 is replaced by $(1,3)$. Note that according to Definition 3, $Share^{G_3}$ and $Take^{G_1}$ are identical as they both raise my payoff by 1 and require the other person to sacrifice a payoff of 1. Hence, we do not expect a strict MAT relation between $A_{Share^{G_3}}$ and $A_{Take^{G_1}}$. Wittig et al. [2013] find that 92% subjects chose Accept in G4; 100% subjects chose Accept

in G3; 100% subjects chose Tolerate in G1; 87% subjects chose Accept in G2.

Cho et al. [2020] run only G2, G3, G4 in Figure 5 across young and old adults. They found that on average approximately 72% subjects chose Accept in G4; 95% subjects choose Accept in G3; 57% subjects chose Accept in G2. Note that, since the paper does not provide the exact percentages, we can only obtain rough estimates from their Figure 2 and Figure 3.

Across these four experiments we report here on ultimatum mini-games, the only inconsistency between the data and Axiom R' is that the acceptance rate in G3 is higher than that in G4, which is clearly driven by fairness concern. We expect that if fairness concern is properly controlled for in a similar ultimatum mini-game experiment, the data should agree with Axiom R'.

Example 2: Stackelberg Mini-Game [Cox et al., 2008]

Consider the following Stackelberg game. The first player chooses an output level x. The second player observes x and chooses an output level q. The price is $p = 30 - x - q$ and both players have constant marginal cost 6 and no fixed cost. Hence, the profit for the second player is $m = (24 - x - q)q$ and the profit for the first player is $y = (24 - x - q)x$.

The Stackelberg mini-game restricts the first player's choices to only two output levels. Cox et al. [2008] consider two of such games, one (G_1) with $q \in \{6, 9\}$ and the other (G_2) with $q \in \{9, 12\}$. Let q^{G_x} denote the opportunity set induces by q chosen in Game $x \in \{1, 2\}$.

Simple calculation shows that $m^*_{6^{G_1}} = 81$, $y^*_{6^{G_1}} = 108$, $m^*_{9^{G_1}} = m^*_{9^{G_2}} = 56.25$, $y^*_{9^{G_1}} = y^*_{9^{G_2}} = 135$, $m^*_{12^{G_1}} = 36$, $y^*_{12^{G_1}} = 144$. By Definition 3, we have 6^{G_1} MGT 9^{G_1} more than 9^{G_2} MGT 12^{G_2}. Then by Axiom R', we expect that $A_{6^{G_1}}$ MAT $A_{9^{G_2}}$. Unfortunately, Cox et al. [2008] only have 5 observations on the choice of 6 in Game 1. Hence, we can not use their experimental data to verify Axiom R'.