On The Role of Sacrifice in Reciprocity*

Simin He[†] and Jiabin Wu[‡]

May 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.

*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 Amanda Chuan, the participants of Warwick-SUFE workshop 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]'s 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.

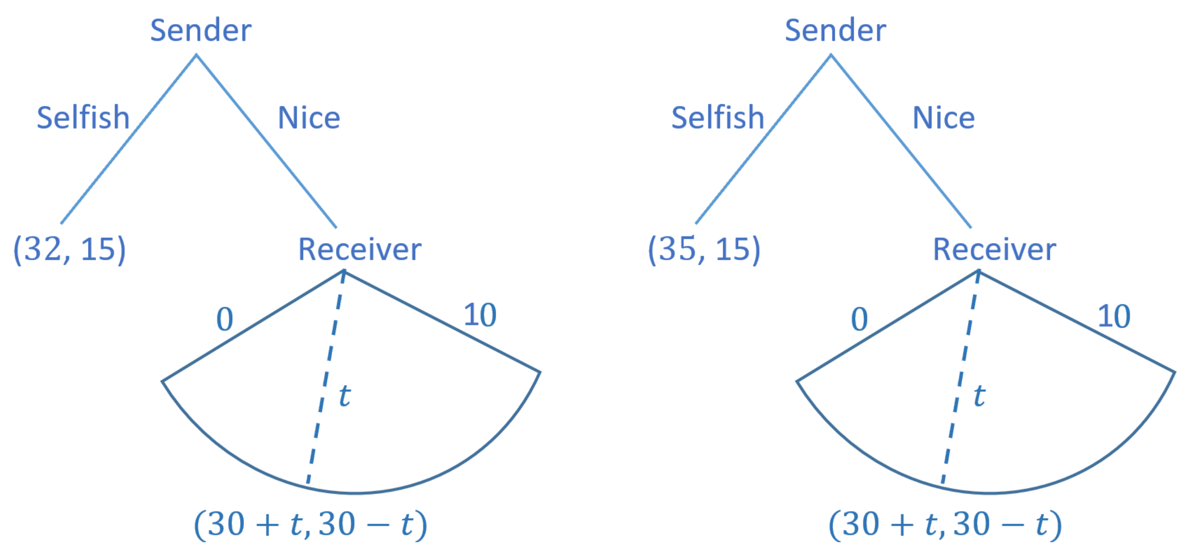

Figure 1: Two Sender-Receiver Games

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

they do not focus on the role of sacrifice in determining perceived intention involved in reciprocal decisions.

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

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 transfer $t \in [0, 10]$ 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 transfer 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 transfer 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 a transfer level. 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 transfer. A sender's willingness to sacrifice is directly revealed by how many times he chooses Nice, and the relationship between the receiver's transfer and this choice frequency —holding the outcome of the selected game fixed— measures the strength of the principle of sacrifice. The relationship between the receiver's transfer and the outcome they experience (15 or 30) —holding fixed the choice profile of the sender— measures the strength of the principle of helping. In addition, we control for fairness concern by making the outcomes of both players equal 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.

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.

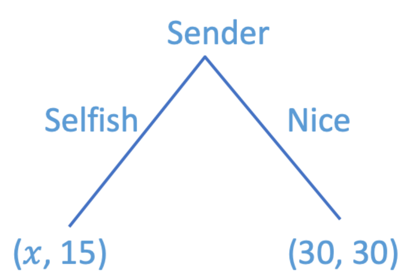

Figure 2: The experimental game

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 ask the receivers to allocate 10 points between themselves and their matched sender. 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 with 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 making both players' outcome equal given the sender chooses the Nice action (30=30) frees the receivers from any serious distributional concern when making their decisions in the allocation task facing the outcome of (30, 30).

2.2 Procedures

The experiment was conducted at the Shanghai University of Finance and Economics in May and June 2021. Chinese subjects were recruited from the subjects pool of the Economic

Lab via Ancademy.[4] We ran 11 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 28. In total, 210 subjects were recruited, most of whom were undergraduate students from various fields of study.[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.

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

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

[5] Detailed summary statistics on the subjects' demographic characteristics are provided in Appendix B.

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

[7] We used the payment system of Ancademy.

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 (weakly) higher level of reciprocity of the receiver. That is, the levels of reciprocity satisfy $c30 \geq b30 \geq a30$, and $b15 \geq a15$.

Hypothesis 2. Conditioned on the same type of sender, a higher outcome of the receiver induces a (weakly) higher level of reciprocity of the receiver. That is, the levels of reciprocity satisfy $a30 \geq a15$, and $b30 \geq b15$.

3 Results

3.1 Types of senders

In total, we have 105 pairs of senders and receivers in our experiment. Among the 105 senders, there are in total four types of choice profiles: 48 senders (45.7%) choose Nice when $x = 28$, and choose Selfish otherwise (type-a, selfish), 20 senders (19.0%) choose Nice when $x = 28$ or 32, and choose Selfish if $x = 35$ (type-b, somewhat nice), and 33 (31.4%) senders choose Nice in all possible values of x (type-c, very nice). Finally, there are 4 senders (3.8%) who choose Selfish for all possible values of x (type-a', anti-social), we exclude them because they are rare.

3.2 Reciprocity level

Table 1 presents the average reciprocity level of the receiver by the sender's type and receiver's realized outcome. We can see that, conditioning on the receiver receiving an outcome of 15, the reciprocity levels are always very low, no matter the sender is selfish

[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.

(type-a) or somewhat nice (type-b). There is no significance between these two cases. Why receivers give back almost nothing to the sender may be attributed to the large payoff inequality because whenever the receiver receives 15, the sender receives at least 28. However, conditioning on the receiver receiving an outcome of 30, the average reciprocity level is 0.73 if the sender is "selfish" (type-a), it increases to 3.75 if the sender is "somewhat nice" (type-b), and it further increases to 4.06 if the sender is "very nice" (type-c). In summary, we find that, conditioning on receiving the same outcome, the reciprocity level weakly increases in the sender's kindness level. This is in support of 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	0.55	0.38	–	$p = 1.000$ (a vs.b)
	(n=33)	(n=8)		
30	0.73	3.75	4.06	$p < 0.001$ (a vs. b)
	(n=15)	(n=12)	(n=33)	$p < 0.001$ (a vs. c)
				$p = 0.147$ (b vs. c)
Diff test	$p = 0.163$	$p < 0.01$		

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. Conditioning on the same outcome, a kinder sender type induces weakly higher reciprocity level of the receiver. That is, reciprocity level $c30 = b30 > a30$, and $b15 = a15$.

Next, we compare the reciprocity level when senders made exactly the same three choices but the receivers receive a different outcome because of chance. First, when the sender is selfish (type-a), we can see that the receivers give back slightly more if they receive 30 instead of 15. However, the difference is insignificant. Second, when the sender is somewhat nice (type-b), the receivers give significantly more to the sender if they receive 30. In summary, we find that, conditioning on facing the same type of senders, the reciprocity level weakly increases in the receiver's payoff. This is in support of Hypothesis 2.

Result 2. Conditioning on the same sender type, higher outcomes of the receivers induce weakly higher reciprocity level of the receiver. That is, reciprocity level $a30 = a15$, and $b30 > b15$.

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. 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 transferred from player 2 to player 1, and $5 = (0 + 10)/2$ is the equitable payoff that is used to measure player 2's kindnesses corresponding to different 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

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

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

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

Player 2's optimal choice is $t = 10$.

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 $x = 35$.[9] Nevertheless, neither has a clear definition on how sacrifice is related to the degree of kindness of a kind action.[10]

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.

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

[9] 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.

[10] 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!

However, this model fails to support our Result 2 (Hypothesis 2) that $b30 \geq b15$ and $a30 \geq a15$ 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)* $m^*_G - m^*_F \geq y^*_G - y^*_F$.

$m^*_G(y^*_G)$ is the highest payoff I (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 the sender's outside option x taking values from $\{28, 32, 35\}$, which essentially corresponds to the three games considered in our experiment.

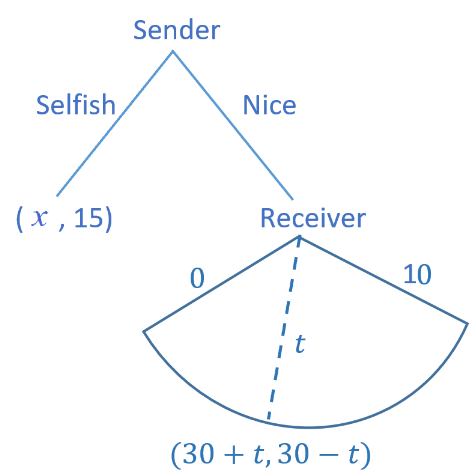

Figure 3: A Sender-Receiver Game

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) $m^*_{\tilde{N}} - m^*_{\tilde{S}} = 30 - 15 = 15$ 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)!

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 $(\tilde{N}, \tilde{S}_{32}, \tilde{S}_{35})$). 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 $y_{\tilde{N}_{35}} \geq y_{\tilde{N}_{32}} \geq y_{\tilde{N}_{28}}$. In 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.

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

We argue that the relevant alternative I need to consider for each set of opportunity sets should be its least generous opportunity set because it gives me the largest monetary payoff possible. Hence, when comparing two opportunity sets that belong to two different sets of opportunity sets respectively, if one gives me more and requires a larger sacrifice from the other comparing to its respective least generous opportunity set than the other does, then I should feel more altruistic given the former than the latter.

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. We finally 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.

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.

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.

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. 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.

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

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.

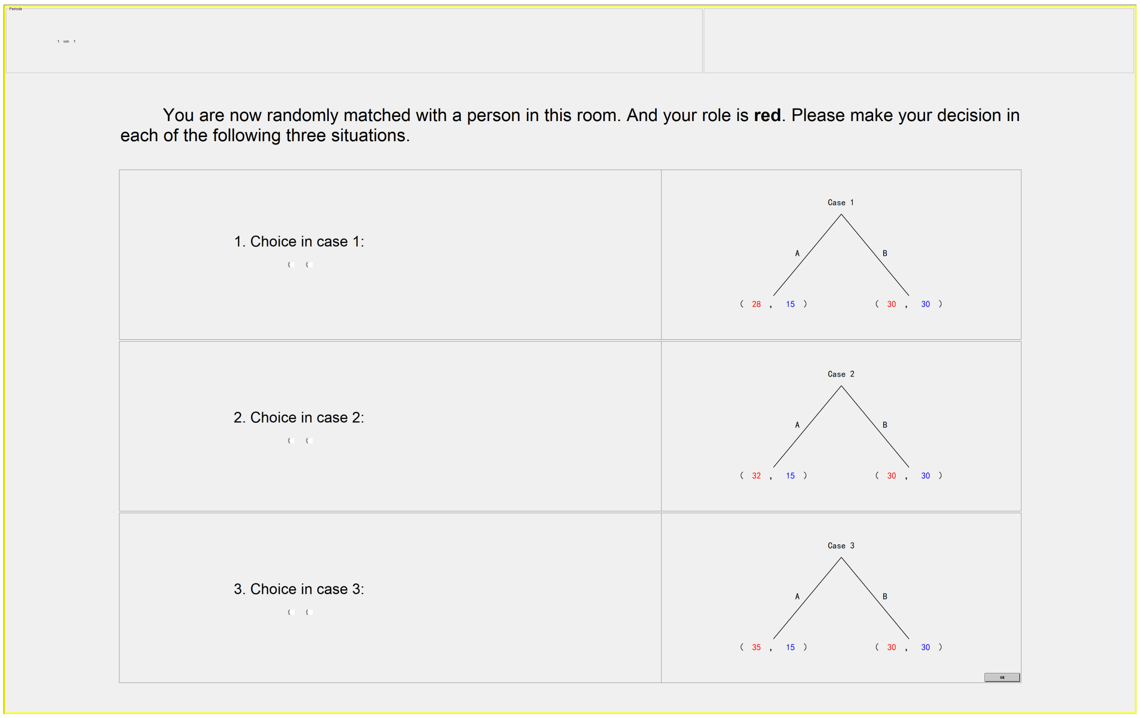

Figure 4: A (translated) screenshot of the experiment

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: Descriptive statistics

Table 2: Demographic information of the senders and the receivers.

	Female, %	Age	Grade	Bussiness/Econ, %
Senders	78.1	20.9	1.95	90.5
	(4.06)	(0.186)	(0.166)	(2.88)
Receivers	74.3	21.0	2.17	87.6
	(4.29)	(0.188)	(0.176)	(3.23)
p-value	0.518	0.671	0.353	0.508

Notes: Standard errors in parentheses. In the last row, *p*-values refer to Kruskal–Wallis tests of equality-of-populations between role types of subjects.