Generalized Reciprocity: Theory and Experiment

Amanda Chuan* Hanzhe Zhang†

August 20, 2021‡

Abstract

Those who have received help may be more likely to help a third party. This pay-it-forward behavior, called generalized reciprocity (or upstream indirect reciprocity), can be powerful in spreading kindness from person to person. We use a series of laboratory games to explore the conditions that support generalized reciprocity. We rationalize our experimental results in a psychological game-theoretic framework that extends Dufwenberg and Kirchsteiger (2004), Battigalli and Dufwenberg (2009), and Fehr and Schmidt (1999). We find that adding altruism, fairness, and reciprocity into the utility function rationalizes our experimental findings. Pay-it-forward behaviors are not motivated by reciprocity motives alone, but also depend on altruism, equity concerns, and higher-order beliefs of others' kindness.

Keywords: reciprocity, pay-it-forward, credit attribution, psychological game theory

JEL: C79, C90, C91

*School of Human Resources and Labor Relations, Michigan State University, achuan@msu.edu.

†Department of Economics, Michigan State University, hanzhe@msu.edu.

‡We thank Tyler Bowen and Kit Zhou for excellent research assistance. Zhang acknowledges the National Science Foundation and Michigan State University Faculty Initiatives Fund for financial support.

1 Introduction

Prior work has shown that people who have received help are more likely to help a third party, even at their own expense (Ben-Ner et al., 2004; Bartlett and DeSteno, 2006; Desteno et al., 2010; Herne et al., 2013; Gray et al., 2014; Tsvetkova and Macy, 2014; van Apeldoorn and Schram, 2016; Mujcic and Leibbrandt, 2018; Simpson et al., 2018; Melamed et al., 2020). This pay-it-forward behavior, which we call *generalized reciprocity* (*upstream indirect reciprocity*), is core to the propagation of kindness from one person to another.[1] At the workplace, an employee who was mentored by a superior may elect to advise a new coworker in turn. In fast food restaurant drive-throughs, when one customer pays for a second customer's order, the second customer may then pay for the next customer after her. Pay-it-forward behavior can transform social norms and create cultures of cooperation (Binmore, 1994; McCullough et al., 2008). These generalized reciprocity chains can grow to include hundreds of people. For example, at a Dairy Queen restaurant drive-through in Minnesota, over 900 consecutive cars chose to pay for the order of the car behind them (Ebrahimji, 2020).

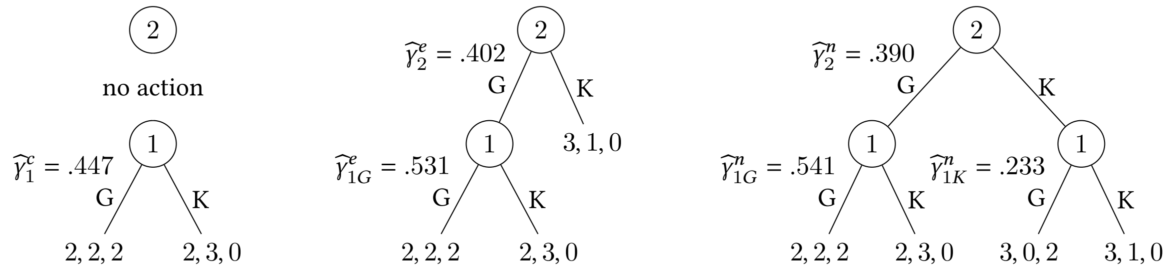
Generalized reciprocity cannot be explained by standard models in which individuals only care about their own material payoffs. Both theoretical and empirical work have shown that psychological components are necessary to explain giving behavior in multiplayer interactions (Rabin, 1993; see Fehr and Gächter, 2000 for an overview). Dufwenberg and Kirchsteiger (2004) develop a workhorse model for dynamic games that involve *direct* reciprocity, in which players can pay back another player's actions. Battigalli and Dufwenberg (2009) present a more general framework of dynamic psychological games and apply it in subsequent work to incorporate emotions, reciprocity, image concerns, and self-esteem in economic analysis.[2] Due to their focus on the strategic nature of direct reciprocity interactions, these models cannot be readily applied to *generalized* reciprocity, in which people cannot pay back their benefactors but must choose whether to pay it forward by benefiting an unrelated third party.

Investigating generalized reciprocity yields unique insights regarding altruistic behavior. In direct reciprocity games, a player may help another because she expects her beneficiary to pay

[1] Under Mujcic and Leibbrandt (2018) and van Apeldoorn and Schram (2016), generalized reciprocity is also called *upstream indirect reciprocity*, in which a player pays forward kindness because she has received kindness in the past. It is not our goal to address *downstream indirect reciprocity*, where a player decides whether to benefit another based on the other player's past actions (Takahashi, 2000; Yoeli et al., 2013; Engelmann and Fischbacher, 2009; Ong and Lin, 2011; Khadjavi, 2017). Upstream indirect reciprocity is much less studied compared to downstream indirect reciprocity, especially in the economics literature; see studies on downstream indirect reciprocity: Bolton et al. (2005); Seinen and Schram (2006); Zeckhauser et al. (2006); Berger (2011); Charness et al. (2011); Heller and Mohlin (2017); Gong and Yang (2019); Gaudeul et al. (2021).

[2] See Battigalli et al. (2019) and Battigalli and Dufwenberg (2021) for recent surveys on theoretical, experimental, applied, and methodological work, and also Falk and Fischbacher (2006) for an alternative but related theory of reciprocity.

Figure 1: Games and giving rates in the experiment



(a) Control **(b)** Exclusive giving **(c)** Nonexclusive giving

Note: A solid line indicates giving away a chip for the gain of two chips for the next player, and a dashed line indicates keeping the chips so that the next player does not gain. Material payoffs are (π_2, π_1, π_0). The $\widehat{\gamma}$s indicate the giving rates in our experiment.

her back. Direct reciprocity games therefore make it difficult to distinguish whether a player helps another due to altruism or to maximize own expected payoffs. In games without direct reciprocity considerations, players know that their beneficiaries cannot pay them back, so any decision to give goes against maximizing their own material payoffs.

To determine the components that create pay-it-forward behavior, we embed altruism, inequity aversion, generalized reciprocity, and higher-order beliefs into a psychological game-theoretic framework that combines and extends Dufwenberg and Kirchsteiger (2004) and Fehr and Schmidt (1999). We test the model's theoretical predictions in a laboratory experiment involving three-player games, depicted in Figure 1. In all games, players choose whether to pass a chip worth \$1 to the closest downstream player. Following the multiplier methods used for investment and public goods games, each chip that is passed turns into two chips for the recipient. Each player's index denotes the number of players behind them in the chain. P0 is the last potential recipient, P1 is the last player to decide on giving, and P2 (if allowed to give) is the penultimate player to decide on giving.

Figure 1a displays the *control game*, in which P2 is endowed with 2 chips and cannot give a chip, P1 is endowed with 3 chips, and P0 with no chip. Only P1 makes a giving decision. In the *treatment games* depicted in Figures 1b and 1c, P2 is endowed with three chips, P1 with one chip, and P0 with no chip. P2 and P1 make giving decisions. P2 can give a chip to P1 so that P1 has 3 chips in total. If P2 gives, all three games have the interim allocation $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$ before P1 makes a giving decision. If P1 keeps, the game concludes with payoffs $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$, and if P1 gives, the game concludes with payoffs $(\pi_2, \pi_1, \pi_0) = (2, 2, 2)$. We thus keep payoff distributions the same in the three games, so that differences in P1's giving behavior across games cannot arise from absolute or relative allocation concerns.

The two treatment games differ in whether P0's channel of receiving is exclusive. In the *exclusive* game depicted by Figure 1b, P1 cannot give P0 a chip unless P2 gives P1 a chip first. If P2 keeps, the game concludes with payoffs $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$. However in the *nonexclusive game* depicted by Figure 1c, P1 can give P0 a chip regardless of whether P2 gives a chip to P1 first. If P1 chooses to keep after P2 keeps, the game concludes with payoffs $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$. If P1 gives even after P2 keeps, the game concludes with payoffs $(\pi_2, \pi_1, \pi_0) = (3, 0, 2)$.[3]

We first compare the giving decisions of last movers (P1 in all three games). If only material payoffs mattered, no one would give. If utility depended on own material payoffs, others' material payoffs, and/or altruism over others' material payoffs, P1 would be equally likely to give in all nodes where she has three chips (the control game and the treatment games after P2 gives). These predictions are inconsistent with our experimental results and provide evidence for the existence of indirect reciprocity, as shown by giving rates in Figure 1. Giving rates $\widehat{\gamma}$ are indexed by a superscript that denotes game type, where c stands for control, e for exclusive, and n for nonexclusive. The subscript G stands for P1's decision after P2 gives, and K for P1's decision after P2 keeps. P1 is (i) most likely to give when P2 has given in the exclusive game ($\widehat{\gamma}^{e}_{1G} = .541$); (ii) second most likely to give when P2 has given in the nonexclusive game ($\widehat{\gamma}^{n}_{1G} = .531$); (iii) third most likely to give when P2 cannot give in control game ($\widehat{\gamma}^{c}_{1} = .447$); and (iv) least likely to give when P2 can give but decides not to in the nonexclusive game ($\widehat{\gamma}^{n}_{1K} = .233$).

We establish the generalized reciprocity effect by comparing P1's actions across games. Compared with control, P1 is $8.4-9.4$ percentage points ($18-21\%$) more likely to give in the exclusive and nonexclusive games after receiving a chip from P2 ($p < 0.005$). The game structure allows us to rule out alternative explanations like income effects, distributional preferences, and social image concerns, which could all motivate P1's giving in the following ways. First, receiving a gift from P2 increases P1's wealth, so P1 may give due to an income effect rather than from reciprocity motives. Second, preferences over payoff distributions may lead subjects to experience disutility if one player has more than another (inequity aversion) or if one player ends the game with nothing (maximin preferences). P1 may be more inclined to give after P2 gives since she could change the payoff distribution from $(2, 3, 0)$ to $(2, 2, 2)$. In contrast, if P2 were to keep, no action on P1's part could equalize payoffs or enable all players to end the game with at least one chip. Lastly, social image concerns may increase the pressure to give after one has received a donation. For these reasons, we keep income, relative payoffs, and social concerns the same between treatment and control. In all games, P1 must choose between $(2, 2, 2)$ if she gives versus $(2, 3, 0)$ if she keeps. Any difference in P1's giving rate between treatment and control must arise from reciprocity motives, since the only difference is whether P1 receives 3 chips from P2's

[3] This distinction could influence P2's impact on P1's and P0's payoffs as well as P1's impact on P0's payoff. We will provide suggestive evidence that differential impact could explain differences in giving behavior across the two treatment games.

generosity or from endowments. Adding reciprocity motives to the model allows us to match theoretical predictions with experimental behaviors.

We then compare the behavior of first movers: P2 in the treatment games and P1 in the control game. P1's giving rate in the control game is $\widehat{\gamma}_1^c = 0.447$, which is significantly greater than P2's giving rates of $\widehat{\gamma}_2^e = 0.402$ and $\widehat{\gamma}_2^n = 0.390$ in the exclusive and nonexclusive games, respectively ($p < 0.05$). As our model will show, inequity aversion is necessary to explain these results. Intuitively, P1 can directly equalize payoffs through giving in the control game, resulting in allocations of $(2, 2, 2)$ rather than $(2, 3, 0)$. However, P2 risks an unequal allocation of $(2, 3, 0)$ by giving in the treatment games since she cannot control what P1 will do after she gives. Models with only altruism and reciprocity would predict the opposite: Since P2's decision affects both P1 and P0, she should be more likely to give than P1 in the control game, who only affects P0's payoff.

Our contributions are threefold. Most game-theoretic frameworks explaining direct reciprocity are limited to two people (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Seinen and Schram, 2006; Cox et al., 2007; Berger, 2011; Gong and Yang, 2019; Gaudeul et al., 2021), but cooperative communities frequently involve interactions among three or more.[4] To the best of our knowledge, our paper is the first to develop a *behavioral game-theoretic* framework for the systematic investigation of how generalized reciprocity occurs, promoting reciprocal exchange beyond two people. The model shows that pure altruism, inequity aversion, and generalized reciprocity incentives are all necessary components to rationalize observed pay-it-forward behavior. Since generalized reciprocity is the channel by which kind actions beget further kind actions, our study constitutes a first step toward understanding how helping behavior can spread from person to person. In doing so, our paper contributes to our understanding of how to foster cultures of cooperation within workplaces, neighborhoods, and communities.

Second, we introduce a simple, novel experiment that establishes the role of generalized reciprocity motives while controlling for alternative explanations. Many prior papers on generalized reciprocity cannot rule out the income effect (Herne et al., 2013; van Apeldoorn and Schram, 2016; Simpson et al., 2018; Mujcic and Leibbrandt, 2018), where the act of receiving a gift itself can make subjects more likely to give, through increasing resources for giving. Furthermore, to the best of our knowledge, our paper is the first to experimentally account for relative wealth differences, which could lead to pay-it-forward behavior if subjects exhibit inequity aversion or minimax preferences. Since our experiment holds social concerns constant across games, we leverage our within-subject design to difference out the impact of social image or social pressure

[4] Some exceptions include Wu (2018) and Jiang and Wu (2019). Reciprocal behavior has also been investigated in other disciplines, including evolutionary biology (Nowak and Sigmund, 1998a,b; Ohtsuki and Iwasa, 2006; Iwagami and Masuda, 2010) and psychology (Hu et al., 2019; Nava et al., 2019).

considerations, which (Charness and Rabin, 2002; Sobel, 2005; Cox et al., 2008) theoretically argue and Malmendier et al. (2014) experimentally find to be important in reciprocal interactions.

Third, our theory and experiment work together to distinguish models of fairness based on outcomes, types, and intentionality. Outcome-based models propose that fairness depends on players' relative payoffs, so inequity aversion and minimax preferences should drive how subjects allocate wealth between themselves and others (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Type-based models posit that giving behavior depends on one's innate altruism parameter (Levine, 1998; Cox et al., 2007; Malmendier et al., 2014). Intentions-based models argue that utility also depends on beliefs about others' kindness intentions (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Battigalli and Dufwenberg, 2009; Gul and Pesendorfer, 2016). As our paper will demonstrate, outcome- and type-based models which ignore intentionality cannot rationalize our experimental findings. By comparing within subjects across games, we show that the same altruism and payoff distribution generate differences in P1's giving between the treatment and control games. Such behavior can only arise if subjects' conceptions of fairness depend on their beliefs of others' intentions, namely if P1's utility depends on her beliefs of P2's kindness.

The rest of the paper is organized as follows. Section 2 introduces the psychological game-theoretic framework, defines the solution concept of *dynamic reciprocity equilibrium*, describes the games, and derives the predicted equilibrium individual giving rates. Section 3 describes the experiment. Section 4 compares our experimental results with theoretical predictions, shows the necessity of altruism, fairness, and reciprocity in rationalizing our results, and suggests future research avenues (e.g., incorporating credit attribution). Section 5 concludes, and the appendices collect omitted proofs and additional experimental results.

2 Theory

2.1 Dynamic reciprocity equilibrium

We consider finite-action multistage games with observable actions and without moves of nature. That is, play proceeds in stages in which each player, along any path reaching that stage, (i) knows all preceding choices, (ii) moves exactly once, and (iii) obtains no information about other players' choices in that stage. Since we extend the direct reciprocity framework of Dufwenberg and Kirchsteiger (2004), we adopt and modify their notation.

Formally, let $N = \{1, \ldots, n\}$ denote the set of players. Let h denote a history of preceding choices represented by a node in the extensive-form representation of games, and let H denote the set of histories of a game. The set of behavioral strategies of player $i \in N$ is denoted by Σ_i, where a strategy $\sigma_i \in \Sigma_i$ of player i assigns a probability distribution over the set of possible choices of player i for each history $h \in H$. Let $\Sigma = \prod_{i \in N} \Sigma_i$ denote the collection of behavioral strategy profiles σ of all players, and $\Sigma_{-i} = \prod_{j \in N \setminus \{i\}} \Sigma_j$ the collection of behavioral strategy profiles σ_{-i} of all players other than i. Let Σ'_{ij} be the set of beliefs of player i about the strategy of player j (i.e., i's first-order beliefs). Let Σ''_{ijk} be the set of beliefs of player i about the belief of player j about the strategy of player k (i.e., i's second-order beliefs). By definition, $\Sigma'_{ij} = \Sigma_j$ and $\Sigma''_{ijk} = \Sigma'_{jk} = \Sigma_k$.

With $\sigma_i \in \Sigma_i$ and $h \in H$, let $\sigma_i(h)$ denote the updated strategy that prescribes the same choices as σ_i, except for the choices that define history h. Note that $\sigma_i(h)$ is uniquely defined for any history h. For any beliefs $\sigma'_{ij} \in \sigma'_{ij}$ or $\sigma''_{ijk} \in \Sigma''_{ijk}$, define updated beliefs $\sigma'_{ij}(h)$ and $\sigma''_{ijk}(h)$ analogously.

Each player's utility function consists of her own material payoff $\pi_i(\cdot)$ and three psychological components: (i) altruistic payoff, (ii) fairness payoff, and (iii) reciprocity payoff. First, player i's material payoff function is $\pi_i : \Sigma \to \mathbb{R}$, which represents money in our experiment. To describe the other components of the utility function, we define players' kindness and beliefs about kindness.[5]

Player j's *equitable payoff* with respect to i is the average between j's lowest and highest possible material payoff based on i's strategy:

$$\pi_j^{Q_i}((\sigma'_{ij})_{j \neq i}) = \frac{1}{2} \cdot \left[\max_{\sigma_i \in \Sigma_i} \pi_j(\sigma_i, (\sigma'_{ij})_{j \neq i}) + \min_{\sigma_i \in \Sigma_i} \pi_j(\sigma_i, (\sigma'_{ij})_{j \neq i}) \right].$$

Note that the equitable payoff is the *expected* material payoff of player j conditional on i's belief about other players' actions.

Player i's *kindness* from playing action $\sigma_i(h)$ to player $j \neq i$ at history $h \in H$ is given by the function $\kappa_{ij} : \Sigma_i \times \prod_{j \neq i} \Sigma'_{ij} \to \mathbb{R}$ defined by

$$\kappa_{ij}\left(\sigma_i(h), (\sigma'_{ij}(h))_{j \neq i}\right) = \pi_j\left(\sigma_i(h), (\sigma'_{ij}(h))_{j \neq i}\right) - \pi_j^{Q_i}\left((\sigma'_{ij}(h))_{j \neq i}\right).$$

Since kindness is defined relative to j's equitable payoff, i's kindness to j is zero if i chooses an action that yields j a material payoff that is j's equitable payoff with respect to i, and i's kindness is positive (negative) if i chooses an action that gives a strictly higher (lower) expected payoff for j than j's equitable payoff.

In general, for any player i, j, and k, player *i's belief about player j's kindness toward player k from taking believed action $\sigma'_{ij}(h)$* at history $h \in H$ is given by the function $\lambda_{ijk} : \Sigma'_{ij} \times \prod_{l \neq j} \Sigma''_{ijl} \to$

[5]In our experiment, all strategies are efficient as defined by Dufwenberg and Kirchsteiger (2004) and Battigalli and Dufwenberg (2009).

$\mathbb{R}$ defined by

$$\lambda_{ijk}(\sigma'_{ij}(h),(\sigma''_{ijl}(h))_{l\neq j}) = \pi_k\left(\sigma'_{ij}(h),(\sigma_{ijl}(h))_{l\neq j}\right) - \pi_k^{Q_j}\left((\sigma''_{ijl}(h))_{l\neq j}\right).$$

Particularly relevant to our games is when $k = i$, where player i's belief about player j's kindness of believed action $\sigma'_{ij}(h)$ toward player i at history $h \in H$ is given by the function $\lambda_{iji} : \Sigma'_{ij} \times \prod_{k\neq j}\Sigma''_{ijk} \to \mathbb{R}$. In the equilibrium we define below, the function λ_{iji} is mathematically equivalent to κ_{ji}. However, they may differ out of equilibrium, since λ_{iji} captures i's beliefs while κ_{ji} captures j's beliefs.

Player i's utility function depends on strategies, first-order beliefs, and second-order beliefs, which we summarize by a vector, $\vec{\sigma} \equiv (\sigma, \sigma', \sigma'')$. These strategies and beliefs in turn determine the expected payoffs of players. The utility function then takes the following form:

$$\begin{aligned} u_i(\vec{\sigma}) = \pi_i(\sigma) + &\underbrace{A_i \sum_{j\neq i} \pi_j(\sigma)}_{\text{altruism}} \\ &- \underbrace{\sum_{s\in S}\sigma(s)\left[\overbrace{\alpha_i \frac{1}{n-1}\sum_{j\neq i}\max\{\pi_j(s)-\pi_i(s),0\}}^{\text{disadvantageous inequity aversion}} + \overbrace{\beta_i \frac{1}{n-1}\sum_{j\neq i}\max\{\pi_i(s)-\pi_j(s),0\}}^{\text{advantageous inequity aversion}}\right]}_{\text{inequity aversion}} \\ &+ \underbrace{\overbrace{\sum_{j\neq i} Y_i \lambda_{iji}(\vec{\sigma})\kappa_{ij}(\vec{\sigma})}^{\text{direct reciprocity}} + \overbrace{\sum_{j,k\neq i} Z_i \lambda_{iki}(\vec{\sigma})\kappa_{ij}(\vec{\sigma})}^{\text{generalized reciprocity}}}_{\text{reciprocity}}, \end{aligned}$$

where $\pi_i(\sigma)$ is i's material payoff; $A_i \in [0,1]$ is i's altruistic factor that dictates how much utility i derives from the material payoffs of other players regardless of the distribution of relative wealth; κ_{ij} is i's kindness to j from choosing strategy σ_i while other players choose σ_{-i}; λ_{iki} is i's belief of k's kindness to i given i's belief of k's belief of other players' strategies; Y_i is i's direct reciprocity parameter; and Z_i is i's generalized reciprocity parameter. From Fehr and Schmidt (1999), we incorporate inequity aversion parameters α and β. Player i receives disutility α_i for each unit of lower payoff than others ("disadvantageous inequity aversion") and disutility β_i for each unit of higher payoff than others ("advantageous inequity aversion"), where $\alpha_i \geq \beta_i$ and $0 \leq \beta_i \leq 1$.

For expositional convenience, we are loose with our arguments in the utility functions. Each utility component depends on a subset of the collection $\vec{\sigma}$ of strategies σ, first-order beliefs σ', and second-order beliefs σ''. In the definitions of us, λs, and κs above, we specified the exact elements of actions and beliefs they depend on, and we will use the generic argument $\vec{\sigma}$ to indicate that they depend on a combination of strategies and beliefs. Notably, material payoffs depend only on strategies, so they are part of models without psychological considerations ("standard" models). Other components all depend on some combination of strategies and beliefs, so they are additional psychological components.

Compared with Dufwenberg and Kirchsteiger (2004), there are three additions. First, we incorporate an altruistic payoff component. Second, we incorporate a generalized reciprocity component. Third, we follow Fehr and Schmidt (1999) and incorporate inequity aversion parameters. By setting $A_i = 0$, $\alpha_i = 0$, $\beta_i = 0$, and $Z_i = 0$ for all i, our model becomes exactly that of Dufwenberg and Kirchsteiger (2004). Since our goal is to explore generalized reciprocity rather than direct reciprocity, subjects do not directly reciprocate in our experiment and we do not estimate direct reciprocity factor Y_i.

Below we define alternative specifications of the utility function in which some psychological components are assumed away. We formulate key predictions based on these alternative utility specifications and test them against our experimental results in Section 4.

Definition 1. ***Altruistic, inequity averse, and reciprocal (AIR) utility*** assumes that $A_i \geq 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, and $Z_i \geq 0$ for all i. ***Standard/selfish (S) utility*** ignores psychological components and assumes that $A_i = 0$, $\alpha_i = 0$, $\beta_i = 0$, $Y_i = 0$, and $Z_i = 0$ for all i. ***Altruistic (A)***, ***Reciprocal (R)***, ***Inequity averse (I)***, ***AI***, ***AR***, and ***IR*** utilities respectively assume the relevant utility components to be nonnegative and other utility components to be zero. See the top rows of Table 1 for the complete parametric specification of each utility function.

Note that the reciprocity utility component depends on strategies, beliefs, and other players' material payoffs. Therefore, the equilibrium is defined with respect to both strategies and beliefs.

Definition 2. Strategies and beliefs $\vec{\sigma}$ constitute a ***dynamic reciprocity equilibrium*** if and only if (i) (consistency) players have correct beliefs about other players' actions, i.e., $\sigma = \sigma_i' = \sigma_{ij}''$ for any players i and j; and (ii) (utility maximization) strategy profile σ maximizes players' utilities at each information set given first-order and second-order beliefs σ_i' and σ_{ij}''.

Theorem 1. *A dynamic reciprocity equilibrium always exists.*

A dynamic reciprocity equilibrium is not necessarily unique. However, we design our three experimental games to have generically unique equilibrium outcomes. That is, except for a measure zero set of parameters, equilibrium strategies are uniquely determined. For a measure zero set of parameters, a player may be indifferent between giving and keeping, and hence any probability of giving can constitute an equilibrium, resulting in multiple equilibria.

2.2 Games

We now describe the three three-player games in our experiment. As the introduction explains, the index of the player indicates the number of downstream players in the giving chain. In explaining our games, we focus on how the AIR model describes trade-offs players experience between giving and keeping. We test predictions when different psychological components of the AIR model are taken out. To simplify the exposition of giving rates, we define the following notation.

Definition 3. For any real number $x \in \mathbb{R}$, define $[\![x]\!]$ to be 1 if x is bigger than 1, x if x is between 0 and 1, and 0 if x is smaller than 0. Mathematically, $[\![x]\!] \equiv \max\{0, \min\{1, x\}\}$.

2.2.1 The control game

First, consider the control game (Figure 1a). P2 is endowed with 2 chips, P1 with 3 chips, and P0 with 0 chips. P2 cannot decide on anything in this game, and exists to keep relative payoffs similar to the treatment games. P1 can either keep all 3 chips so that P0 has 0 chips, or pass 1 chip to P0 so that P0 has 2 chips. P0 cannot decide on anything, and can only receive chips from P1.

Lemma 1. *In the control game, P1 prefers giving if and only if* $2 \cdot A_1 + 2 \cdot \beta_1 \geq 1$.

P1 can either keep 1 chip (so that P0 gain no chip) or give up 1 chip (so that P0 gains 2 chips). When P1 gives up one chip, the material payoffs change from $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$ to $(\pi_2, \pi_1, \pi_0) = (2, 2, 2)$. By giving, she lowers her material payoff by 1 unit, but increases her altruistic payoff by $2 \cdot A_1$ units as P0's payoff increases from 0 to 2. Moreover, because inequity aversion gives P1 disutility from having more chips than other players, she gains $2 \cdot \beta$ from giving and equalizing payoffs across all three players. Overall, the psychological gain of giving by P1 is

$$2 \cdot A_1 + 2 \cdot \beta_1. \tag{1}$$

Figure 2a depicts P1's equilibrium giving rate as A_1 varies. In equilibrium, there is no mixed strategy except for when $2 \cdot A_1 + 2 \cdot \beta_1 = 1$, so the equilibrium giving rate can be represented by an indicator function: $\gamma_1^c = 1_{2 \cdot A_1 + 2 \cdot \beta_1 \geq 1}$.[6] P1 is more inclined to give the higher her altruistic factor A_1 and the higher her advantageous inequity aversion β_1 (that is, the more she dislikes having more than other players). Pure altruism and/or advantageous inequity aversion—but not

[6] When $2 \cdot A_1 + 2 \cdot \beta_1 = 1$, P1 is indifferent between giving and keeping, since the change in material payoffs balances out the change in psychological payoffs. In equilibrium, P1 can choose to give with any probability $\gamma_1^c \in [0, 1]$. Without loss of generality, we assume that P1 chooses to give.

disadvantageous inequity aversion or reciprocity—helps rationalize giving by P1 in the control game.

2.2.2 Exclusive game

Second, consider a three-player game in which P0's channel of receiving chips is exclusive (Figure 1b). P2 is endowed with 3 chips, P1 with 1 chip, and P0 with 0. P2 can either keep all 3 chips or give 1 chip to P1 so that P1's chip count increases from 1 to 3. Only upon receiving additional chips can P1 choose to give. If P1 gives 1 chip, P0 gets 2 chips. If P1 keeps, P2 gets 0.

Lemma 2. *In the exclusive game, P2 prefers giving if* $1 + (1 - \gamma^e_{1G}) \cdot \alpha_2/2 \leq (2 + \gamma^e_{1G}) \cdot A_2 + (3/2 + \gamma^e_{1G}) \cdot \beta_2$*, and P1 gives with probability* $\gamma^e_{1G} = [\![(2 \cdot A_1 + 2 \cdot \beta_1 - 1)/Z_1 + 2]\!]$.

Compared to keeping, giving lowers P2's material payoff but increases her utility from altruism and from having an equitable distribution of payoffs among all players in the group. The left-hand side of the inequality in Lemma 2 represents the two ways P2 loses from giving. Compared to keeping, giving lowers P2's material payoff by 1. P2 may also suffer utility loss from disadvantageous inequity, since P1 may keep after she gives, making her material payoff lower than P1's. More precisely, if P1 keeps after P2 gives, P2 incurs utility loss from disadvantageous inequity of $(3 - 2) \cdot \alpha_2/2 = \alpha_2/2$. Because P1 keeps with probability $1 - \gamma^e_{1G}$ after P2 gives, giving would lower P2's expected utility by $(1 - \gamma^e_{1G}) \cdot \alpha_2/2$.

The right-hand side of the inequality represents the two ways P2 gains from giving. Giving increases P2's altruistic payoff by $A_2 \cdot (2 + \gamma^e_{1G})$ in expectation. Giving also rectifies inequality aversion, in that P2 is less likely to have more than other players. If P2 keeps, she suffers disutility of $[(3 - 1) + (3 - 0)] \cdot \beta_2/2 = 5 \cdot \beta_2/2$ from advantageous inequity, since she will have higher payoffs compared to P1 and P0. If P2 gives, two scenarios can happen. If P1 gives, P2 suffers no inequity aversion since all players will have 2 chips. If P1 keeps (which happens with probability $1 - \gamma^e_{1G}$), then P2 suffers $(2 - 0) \cdot \beta_2/2 = \beta_2$ from getting more than P0. Hence, by giving, P2 gains in expectation $5 \cdot \beta_2/2 - (1 - \gamma^e_{1G}) \cdot \beta_2 = (3/2 + \gamma^e_{1G}) \cdot \beta_2$. In summary, altruism and inequity aversion motivate P2 to give.

Upon receiving 2 chips from P2, P1 faces the following trade-off. If P1 gives, P1 loses one unit of material payoff, but gains in the three psychological components. A 2-chip gain for P0 gives P1 an altruistic payoff gain of $2 \cdot A_1$ and $2 \cdot \beta_1$ from equalizing payoffs. Furthermore, P1's kind action of giving earns P1 a generalized reciprocity payoff of $Z_1 \cdot (2 - \gamma''_{1G})$, where γ''_{1G} is P1's belief of P2's belief of P1's probability of giving. Altogether, P1's psychological gain from giving after P2 gives is

$$2 \cdot A_1 + 2 \cdot \beta_1 + Z_1 \cdot (2 - \gamma''_{1G}). \tag{2}$$

In equilibrium, P1's second-order belief must equate with her strategy ($\gamma''_{1G} = \gamma^e_{1G}$). Hence, if $2 \cdot A_1 + 2 \cdot \beta_1 + Z_1 \geq 1$, then $\gamma^e_{1G} = 1$; if $2 \cdot A_1 + 2 \cdot \beta_1 + 2 \cdot Z_1 \leq 1$, then $\gamma^e_{1G} = 0$. Otherwise, P1 gives with a probability strictly between 0 and 1 that makes her indifferent between giving and keeping: $\gamma^e_{1G} = (2 \cdot A_1 + 2 \cdot \beta_1 - 1)/Z_1 + 2$. P1's inclination to give increases with altruism A_1, advantageous inequity aversion β_1, and generalized reciprocity Z_1. Figure 2b depicts how P1's equilibrium giving rate in the exclusive game varies with altruism A_1.

2.2.3 Nonexclusive game

Finally, consider a three-player game in which P0's channel of receiving chips is nonexclusive (Figure 1c). P2 is endowed with 3 chips, P1 with 1, and P0 with 0. P2 can either keep all the chips so that P1's chip count remains unchanged, or give away 1 chip so that P1's chip count increases by 2. Regardless of P2's decision, P1 can keep all the chips or give away 1 chip so that P0's chip count increases by 2.

Lemma 3. *In the nonexclusive game, P2 prefers giving if and only if* $1 + (1/2 - \gamma^n_{1G}/2) \cdot \alpha_2 \leq (2+\gamma^n_{1G}-\gamma^n_{1K}) \cdot A_2 + (3/2+\gamma^n_{1G}-\gamma^n_{1K}/2) \cdot \beta_2$, *P1 gives with probability* $\gamma^n_{1G} = [\![(2 \cdot A_1 + 2 \cdot \beta_1 - 1)/Z_1 + 2]\!]$ *after P2 gives, and P1 gives with probability* $\gamma^n_{1K} = [\![(2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - 1)/Z_1 - 1]\!]$ *after P2 keeps.*

P2's gains from giving are similar to those in the exclusive game. The only difference is that after P2 keeps, P1 gives 1 chip with probability γ^n_{1K}, and 1 new chip gets created from P1's gift to P0. P2 then gains A_2 from altruism and $\beta_2/2$ since payoffs become more equal after P0 gains 2 chips. Since the psychological penalty to keeping is less severe for P2 in the nonexclusive game, the net benefit of giving is smaller in the nonexclusive game than the exclusive game by $(A_2 + \beta_2/2) \cdot \gamma^n_{1K}$.

If P2 chooses to give, then P1 faces the same trade-off as in the exclusive game. Therefore, the equilibrium giving rate γ^n_{1G} in the nonexclusive game is characterized in the same way as in the exclusive game. If P2 chooses to keep, reciprocity motives will make P1 more inclined to keep. If P1 gives instead, her reciprocity motives will generate a utility loss of $Z_1 \cdot (2 - \gamma''_{1K})$. In addition, by giving, P1 changes the material payoffs from $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$ to $(3, 0, 2)$, which results in $3 \cdot \alpha_1/2$ units increase in disadvantageous inequity aversion and β_1 units of increase in advantageous inequity aversion.Overall, the psychological gain of giving by P1 after P2 keeps is

$$2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - Z_1 \cdot (2 - \gamma''_{1G}). \tag{3}$$

Figure 2c depicts P1's giving rates in the nonexclusive game. If P1 gives after P2 keeps, then her decision can only be justified by altruism. Note that there is no simultaneous mixing of both giving decisions in equilibrium under any combination of parameters. This is because the conditions for indifference are different at the two decision nodes, when P1 is indifferent between giving and keeping at one decision node, she is not indifferent at the other.

2.3 Predictions

In this subsection, we predict giving rates under the general AIR utility ($A_i > 0$, $\alpha_i > 0$, $\beta_i > 0$, and $Z_i > 0$). Table 1 summarizes the predictions under different utility functions. Our results center on giving rates comparisons, and we say that one is more inclined to give than another in the following sense.

Definition 4. Player i is ***more (less) inclined*** to take action s at node H than player j to take action s' at node H', $\sigma_i(s|H) \succ$ (resp., $\prec$) $\sigma_j(s'|H')$, if for $\alpha_i = \alpha_j$ and $Z_i = Z_j$, $\sigma_i(s|H) \geq$ (resp., $\leq$) $\sigma_j(s'|H')$ for all parameters, and the inequality holds strictly for some parameters.

2.3.1 Giving by last mover P1

P1 is the last mover in all three games. Figure 3 shows P1's equilibrium giving rates for different altruistic factors. The comparisons are unambiguous: $\gamma^e_{1G} \sim \gamma^n_{1G} \succ \gamma^c_1 \succ \gamma^n_{1K}$. We discuss the pairwise comparisons of these giving decisions in the following five propositions.[7]

First, compare P1's two giving decisions in the nonexclusive game, after P2 gives versus after P2 keeps. Regardless of P2's choice, P1 incurs the same material loss (1 unit) and altruistic gain (2 units) from giving. However, if P1 chooses to give after P2 keeps, she incurs a larger inequity aversion loss and reciprocity loss than if she chooses to give after P2 gives.

Proposition 1. *In the nonexclusive game, P1 is more inclined to give after P2 gives than after P2 keeps. That is,* $\gamma^n_{1G} \succ \gamma^n_{1K}$.

Since there is no simultaneous mixed strategy in equilibrium (as argued in the previous section), Proposition 1 implies that when P1 chooses to give after P2 keeps, P1 must also give after P2 gives. When P1 chooses to keep after P2 gives, P1 must also keep after P2 keeps. It is possible that P1 gives after P2 gives and keeps after P2 keeps, but never possible for P1 to keep after P2 gives and give after P2 keeps. Any subject who does so violates theoretical predictions as long as subjects exhibit reciprocity motives or inequity aversion.

Now compare the psychological gain of giving in the control game to that in the exclusive game. The choice for P1 is the same in terms of material payoffs: either $(2, 2, 2)$ by giving or

[7] There are $4 \times 3/2 = 6$ different pairwise comparisons for the four decisions. We do not directly compare γ^e_{1G} and γ^n_{1K}, but discuss the other five pairwise comparisons in the five propositions.

$(2, 3, 0)$ by keeping. If subjects have reciprocity motives, a gift from P2 increases P1's giving rate in the exclusive game.

Proposition 2. *Relative to the control game, P1 is more inclined to give in the exclusive game after P2 gives. That is,* $\gamma^e_{1G} \succ \gamma^c_1$.

Similarly, a gift from P2 increases P1's giving rate in the nonexclusive game relative to the control game.

Proposition 3. *Relative to the control game, P1 is more inclined to give in the nonexclusive game after P2 gives. That is,* $\gamma^n_{1G} \succ \gamma^c_1$.

However, both reciprocity motives and inequity aversion would decrease P1's giving inclination after P2 keeps in the nonexclusive game.

Proposition 4. *Relative to the control game, P1 is less inclined to give in the nonexclusive game after P2 keeps. That is,* $\gamma^n_{1K} \prec \gamma^c_1$.

Finally, P1's inclination to give is the same in the exclusive and nonexclusive games after P2 gives, and this result holds for all utility preferences we consider.

Proposition 5. *P1 is equally inclined to give after P2 gives in the nonexclusive game and the exclusive game. That is,* $\gamma^n_{1G} \sim \gamma^e_{1G}$.

Note that it is possible that a player is indifferent between giving and keeping in equilibrium in the exclusive and nonexclusive games, because she mixes between giving and keeping in equilibrium. Hence, when subjects are observed to give in one game and keep in another, the difference in observed behavior neither validates nor invalidates the prediction that they are equally likely to give. We discuss this prediction further in Section 4.6.

2.3.2 Giving by initial movers

We compare the first movers in these games: P1 in the control game and P2 in the treatment games. To summarize, the initial mover's giving rate is higher in the exclusive game than in the nonexclusive game, but it is unclear whether the initial mover is more inclined to give in the control game than in the treatment games, since altruism pushes for greater giving while inequity aversion pushes for lower giving in the treatment games.

First, compare P2's choices in the two treatment games. 2 has a lower incentive to give in the nonexclusive game than the exclusive game. In the exclusive game, P2 knows that keeping will prevent P1 from giving, while in the nonexclusive game P2 knows that P1 can still give even if she kept. In particular, knowing that P1 can give even after P2 keeps in the nonexclusive game will increase P2's expected utility from keeping by $\gamma_{1K}^n \cdot A_2$ from altruism and $\gamma_{1K}^n \cdot \beta_2/2$ from having more equal payoffs.Figure 4 depicts giving rates comparison.

Proposition 6. *P2 in the exclusive game is more inclined to give than P2 in the nonexclusive game. That is,* $\gamma_2^e > \gamma_2^n$.

Next, compare the giving rates of P1 in the control game and P2 in the exclusive game. Since we compare the giving decisions for the same subject, we can assume that all psychological parameters are the same for the subject across games and player roles: $A_1 = A_2 \equiv A$, $\alpha_1 = \alpha_2 \equiv \alpha$, $\beta_1 = \beta_2 \equiv \beta$, and $Z_1 = Z_2 \equiv Z$. By giving, P1 in the control game gets an altruistic payoff of $2 \cdot A$, and P2 in the exclusive game gets an altruistic payoff of $(2 + \gamma_{1G}^e) \cdot A$, because P1 in the exclusive game generates additional $\gamma_{1G}^e \cdot A$ units of altruistic payoff for P2 by passing to P0. Therefore, by altruism alone, P1 in the control game would be less inclined to give than P2 in the exclusive game.

Inequity aversion could create the opposite effect from altruism. P2 does not have a sure chance of equalizing payoffs in the exclusive game since P1 may choose to keep after P2 gives, while in the control game P1 will certainly equalize payoffs by choosing to give. This uncertainty decreases expected utility from giving of P2 in the exclusive game compared to P1 in the control game. Furthermore, if P1 keeps, P2's inequity aversion causes her to suffer disutility $(3-2)\cdot\alpha/2 = \alpha/2$ from having a lower payoff than P1. Since this occurs with probability $1 - \gamma_{1G}^e$ and results in material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$, the loss in expectation is $(1 - \gamma_{1G}^e) \cdot \alpha/2$.

Proposition 7. *P1 in the control game is more inclined to give than P2 in the exclusive game, i.e.,* $\gamma_1^c > \gamma_2^e$ *if and only if* $(1 - \gamma_{1G}^e) \cdot \alpha/2 \geq \gamma_{1G}^e \cdot A + (\gamma_{1G}^e - 1/2) \cdot \beta$.

The comparison between P1 in the control game and P2 in the nonexclusive game is similar to the logic above. A higher altruistic payoff incentivizes P2 in the nonexclusive game to give more, since her gift could improve payoffs for both P1 and P0. But, P1 can equalize payoffs with certainty in the control game whereas P2 cannot in the nonexclusive game. In summary, altruism alone would drive greater giving by P2 in the nonexclusive game, but inequity aversion would drive greater giving by P1 in the control game.

Proposition 8. *P1 in the control game is more inclined to give than P2 in the nonexclusive game, i.e.,* $\gamma_1^c > \gamma_2^n$ *if and only if* $(1/2 - \gamma_{1G}^n/2) \cdot \alpha \geq (\gamma_{1G}^n - \gamma_{1K}^n) \cdot A + (\gamma_{1G}^n - \gamma_{1K}^n - 1/2) \cdot \beta$.

2.3.3 Summary of predictions

Table 1 summarizes the predictions under each of the utility functions listed in Definition 1, which turn on or off each of the three psychological components we consider. In the most general AIR utility specification (Column 8), which we elaborated on in previous sections, the giving rates by the last mover are ordered: $\gamma^e_{1G} \sim \gamma^n_{1G} \succ \gamma^c_1 \succ \gamma^n_{1K}$ (Predictions 1-5 in Column 8). The initial mover's giving rate is higher in the exclusive game than the nonexclusive game (Prediction 6 in Column 8), but it is unclear whether the initial mover in the control game is more inclined to give than in the treatment games, since altruism pushes for greater giving and inequity aversion pushes for lower giving in the treatment games (Predictions 7 and 8 in Column 8).

A standard model without the aforementioned psychological components predicts no giving under any circumstance (Column 1). We also consider the predictions with alternative utility specifications in which only one psychological component is considered at a time (Columns 2-4). Under altruism alone (Column 2), there would be a positive giving rate for each of the six decisions, but no distinction in the likelihood of giving for last movers (P1) in different games (Predictions 1-5 in Column 2). Initial movers would be most likely to give in the exclusive game, and equally likely to give in the nonexclusive and control game (Predictions 6-8 in Column 2). Under inequity aversion alone (Column 3), P1's giving rates after P2 gives in the treatment games would equal that in the control game (Predictions 2, 3, and 5 in Column 3), and P1 would not give in the nonexclusive game after P2 keeps (Predictions 1 and 4 in Column 3). There are ambiguous predictions in the comparison of first movers in the control versus treatment games (Predictions 7-8 in Column 3). Under reciprocity motives alone (Column 4), P1 would only give after P2 gives and not after P2 keeps in the nonexclusive game (Prediction 4 in Column 3). This would make the nonexclusive and exclusive games effectively the same to P2, so P2 would be equally inclined to give in the two games (Prediction 6 in Column 3).

We also consider the predictions when one psychological component is omitted (Columns 5-7). Without reciprocity (Column 5), P1 would be equally inclined to give in the control game and after P2 gives in the treatment games (Predictions 6-8 in Column 5 vs 8). Without inequity aversion (Column 6), predictions for the giving rate by last mover are the same as in the general AIR model, but the initial mover would be less inclined to give in the control game than in the treatment games (Predictions 7 and 8 in Column 6). Finally, without altruism (Column 7), the predictions are the same as in the general model, but there would be no reason for P1 to give after P2 keeps in the nonexclusive game (Predictions 1 and 4 in Column 7).

3 Experimental procedure and data

Experimental sessions were implemented on Amazon Mechanical Turk (MTurk), an online platform that is commonly used by experimental social scientists to collect information about behaviors, attitudes, and opinions. The study was administered between February 23 and March 26 of

2021.[8]

We conducted the experiment using Qualtrics software. Since our study involved three-player games, we held sessions of 9 subjects each. Recruitment materials informed subjects that they would receive $3 for completing the study and up to $5 in bonus payments. Subjects could preview all experimental materials before choosing to participate. Experimental materials informed subjects that upon study completion, they would be randomly assigned to a game and a group with other players from their session. Their bonus earnings were calculated based on their giving decisions, as well as the giving decisions of their group mates. Subjects received their base and bonus payments via MTurk within 24 hours of study completion.

It was difficult for subjects to externally manipulate the study, since they made their giving decisions after being told how they would be compensated but before they knew which game, group, or player role they would be compensated for. There was also no way subjects could contact each other, or know whom they would be playing with when they made their decisions. The only way subjects could maximize their utility was to make giving choices that aligned with their true preferences.

We used the strategy method to elicit subjects' actions at all nodes of each game. For example, in the nonexclusive game we asked subjects whether they would give as P1 in the case that P2 gave and whether they would give in the case that P2 kept. We leverage this within-subject variation to examine how each subject's actions differ across player roles and across games. Our goal was to collect three sets of data: giving decisions, attributions of credit for each player in each game, and beliefs about other players' giving strategies. We assess the giving data against theoretical predictions to identify the psychological components necessary for generalized reciprocity. When theoretical predictions are ambiguous, we use data on credit and beliefs about others' giving strategies to refine our predictions.

All subjects proceeded through the experiment as follows. Screenshots of the experiment are shown in Appendix B. First, subjects were taken to the consent page, which described their expectations and rights as study participants. Next, they viewed a video that described the study and all games.[9] Prior to the beginning of each game, subjects viewed the extensive-form diagram of the game they were about to play, which contained information about endowments and payoffs for each realization of the game. On almost every screen, subjects could click on a link that displayed, in a separate tab, the extensive-form diagram and video describing the relevant game.

[8]The experiment was conducted online since it took place during the Covid pandemic, and nonessential in-person studies were prohibited by Michigan State University. The study was approved by the Institutional Review Board at Michigan State University.

[9]The experiment involved three three-player games and two four-player games. This paper compares giving results between the three-player games, since our goal is to develop a framework to explain pay-it-forward behavior. We plan to explore the predictions of the theory on giving results in the four-player games in subsequent studies.

To incentivize subjects to pay attention during the study, we asked quality check questions before the exclusive and nonexclusive games. Subjects were informed they would earn an additional $0.50 per game if they answered all questions about the game correctly on their first attempt.[10]

After the giving questions, we asked subjects about perceived impact, or credit attribution. Subjects viewed a screen that asked, “What percent of X’s payoff is due to Y?” where X and Y each took values of P0, P1, and P2. The sum of credit over all Y had to equal 100% before subjects could proceed. These responses provide suggestive evidence that perceived impact can account for differences in P1’s and P2’s giving behavior in the exclusive versus nonexclusive games (see Section 4.

For the exclusive and nonexclusive games, we also asked subjects about their beliefs regarding other players’ strategies. We were specifically interested in P2’s belief about P1’s likelihood of giving in the exclusive and nonexclusive games, which informs P2’s giving strategy under our model. From the perspective of P2, subjects stated their beliefs of the likelihood that P1 would give to P0. We use these questions as a sanity check that beliefs align with actual giving behavior.

All subjects played the control game first. After the control game, the order of the exclusive and nonexclusive games was randomized. Robustness checks in Section 4.4 show that almost all our results hold independent of game order. At the end of the study, subjects completed a demographic questionnaire.

Table 2 displays summary statistics. The first column shows results for the full sample of 403 subjects.[11] The second column excludes subjects who demonstrably struggled to understand the game and answered three or four out of four quality-check questions wrong. We are left with 324 subjects who answered fewer than two questions wrong, which we call “accurate responders.” The third column excludes subjects who violate Proposition 1, and therefore exhibit behaviors that are inconsistent with any model involving reciprocity motives or inequity aversion. We call this subsample of 364 subjects the “Proposition 1 compliers.”[12]

Overall, the three samples are very similar in terms of demographics. Less than a third of subjects (26.4-27.5%) are women, over 80% have at least an associate’s degree, and over 90% are employed full- or part-time. Around 67-68% are US citizens and US residents, and over 75% are

[10]In addition to the three player games, subjects concluded the study with two four player games that are not the focus of this paper. Subjects could earn $0.50 each for four sets of quality check questions. They could earn $2 maximum on the quality check questions and $3 maximum from their giving decision, so their maximum bonus earnings were $5.

[11]Our full sample is not a multiple of 9 since two subjects in a session did not answer all questions and were omitted from the analysis.

[12]We define these samples in order to identify subjects with consistent, rationalizable behavior across games, so considerable overlap between the two samples would be expected. Of the 403 subjects in the full sample, 302 are in both the accurate responder and the Proposition 1 complier samples. 17 are in neither sample. 22 are in the accurate responder sample but not the Proposition 1 complier sample. 62 are in the Proposition 1 complier sample but not the accurate responder sample.

native English speakers. In terms of race and ethnicity, about half of subjects are white, 30% are Asian, 10-13% are Black, 5% are Hispanic, and the remaining 3% are categorized as other race or ethnicity. In terms of age, half of the subjects are 26-35 years old, 22-23% are 36-45 years old, 15-16% are 16-25 years old, and about 10% of subjects are 46-65 years old. Only 1.2-1.5% of subjects are 65 or older.

The bottom of Table 2 summarizes how subjects performed. Bonus payments ranged from \$0 to \$5, with a median payment of \$3 and an average payment of \$2.55-\$2.73. Subjects were asked four quality-check questions (two each for the exclusive and nonexclusive game). In the full sample, the average number of questions answered incorrectly was 2.16, with a median of 1 question. However, the distribution is positively skewed: Of our 403 total subjects, 140 had no incorrect questions, 87 had one incorrect question, and 97 had two incorrect questions. The remaining 78 had 3 or more incorrect questions. It is likely that those who answered more than two questions incorrectly were not paying attention or had difficulty understanding the games, so their choices may not reflect their true preferences. Below, we present results for the full sample ($N = 403$), the accurate responders sample ($N = 324$), and the Proposition 1 compliers sample ($N = 364$).

Subjects took 28-29 minutes on average to complete the study. Although the study was designed to be completed within an hour, two subjects took 69.53 and 124.90 minutes, respectively. Twenty-eight subjects took between 45 and 60 minutes. Excluding these 30 subjects from the two samples does not appreciably change results for any of our three samples.

4 Results

We first assess the results of our experiment against our eight propositions using the paired one-tailed t-test results summarized in Table 3. Panel a uses the full sample of 403 subjects, panel b restricts the sample to accurate responders, and panel c reports results for Proposition 1 compliers.

First, we compare the actions of last movers (P1) across games (Section 4.1). Comparing the actions of last movers establishes the generalized reciprocity effect and points to the need for both altruism and reciprocity incentives to explain giving behavior. We then compare the actions of first movers (P1 in the control game and P2 in the treatment games) (Section 4.2). Comparing the actions of first movers establishes inequity aversion as a necessary psychological component to align theoretical predictions with experimental results. To establish the robustness of the results, we examine different subsamples who may have more accurate responses (Section 4.4) and we present an alternative way of measuring the theoretical predictions of the model by tabulating the number of individuals who violate the predicted behavior (Section 4.5). Furthermore, a closer examination of the sample (i.e., focusing on the accurate responders sample or the Proposition 1 compliers sample) reveals the role of perceived impact and credit-based reciprocity (discussed in Section 4.6).

4.1 Giving rates by last movers

Proposition 1 compares P1's giving in the nonexclusive game after P2 gives versus after P2 keeps. Across the three samples, P1's giving likelihoods are 54.1%-59.9% after P2 gives and 15.1-23.3% after P2 keeps, with the difference significant at the 0.01% level. This behavior is consistent with models that account for altruism and inequity aversion (AI), altruism and reciprocity (AR), and altruism, inequity aversion, and reciprocity (AIR). Altruism is necessary to explain why P1 has a positive probability of giving even when P2 keeps ($\widehat{\gamma}^n_{1K} > 0$), but is alone insufficient to explain why P1 would give more after P2 gave than after P2 kept ($\widehat{\gamma}^n_{1G} > \widehat{\gamma}^n_{1K}$). Altruism with reciprocity incentives provides one explanation for this behavior, since reciprocity incentives make P1 more likely to give after P2 gave and to keep after P2 kept. Another explanation is that inequity aversion leads P1 to experience disutility from unequal payoffs. After P2 gives, P1 can equalize payoffs by giving to P0 such that all players earn \$2. If P2 kept, however, nothing P1 chooses will equalize payoffs. P1 would therefore be more likely to give after P2 gave than after P2 kept. Proposition 1 shows the necessary role of altruism in explaining P1's behavior in the nonexclusive game, but cannot distinguish whether reciprocity incentives, inequity aversion, or both contribute to why P1's giving likelihood is greater after P2 gave than after P2 kept.

Propositions 2 and 3 establish the generalized reciprocity effect by comparing P1's behavior in the control game versus the treatment games after P2 gave. We find that P1's giving rates are 52.2-53.1% in the exclusive game and 54.1-59.7% in the nonexclusive game after P2 gave. Both values are significantly greater than P1's giving rate of 44.7-45.5% in the control game ($p < 0.01$ across all three samples). The pattern of giving establishes that reciprocity incentives are necessary in explaining behavior in our games, since it contradicts the predictions of all models that exclude reciprocity. In all cases, P1 chooses between payoffs of (2,2,2) if she were to give and (2,3,0) if she were to keep. The game design holds constant the number of players behind P1, P1's own income, and the relative payoffs across all players. The only difference between the treatment and control conditions is that P1's endowment is partly attributable to P2's kindness, rather than experimental conditions. Receiving the gift increases P1's giving likelihood by 7-14 percentage points, indicating that benefiting from another person's kindness made subjects more likely to pay it forward.

Proposition 4 compares P1's giving in the control game with P1's giving in the nonexclusive game if P2 were to keep. Empirical giving rates are 20 percentage points higher in the control game (44.7-45.6%) than in the nonexclusive game case after P2 kept (15.1-23.3%, $p < 0.0001$ across all three samples). This result aligns with all models that incorporate either reciprocity or inequity aversion. Under the reciprocity explanation, like actions beget like actions, so P1 is more inclined to keep her chip if P2 kept hers. Under the inequity aversion explanation, P1 can equalize payoffs in the control game by giving her chip to P0 but not in the nonexclusive game if P2 kept her chip. By comparing results with Proposition 4 predictions, we conclude that reciprocity or inequity aversion is sufficient to align our predictions with P1's behavior.

Proposition 5 examines how P1's behavior differs in the exclusive and nonexclusive games in the case in which P2 gave. Results differ between the three samples. In the full sample, giving rates are 54.1% in the nonexclusive game and 53.1% in the exclusive game. The difference is insignificant, which aligns with the predictions of all models we consider. By itself, Proposition 5 cannot rule out any of our competing models when examining the full sample.

However, differences are significant in the accurate responders sample and Proposition 1 compliers sample. In the former sample, giving rates are 56.2% in the nonexclusive game and 52.2% in the exclusive game ($p < 0.10$). In the latter sample, giving rates are 59.7% in the nonexclusive and 52.6% in the exclusive game ($p < 0.005$).[13] This suggests that an additional psychological component beyond altruism, reciprocity, and inequity aversion may be needed to explain P1's giving behavior in the exclusive versus nonexclusive games. We speculate as to what this component could be and provide suggestive evidence for our conjecture in subsection 4.6.

4.2 Giving rates by first movers

We next turn to examining first movers' decisions by comparing P2 in the treatment games and P1 in the control game. Proposition 6 compares P2's giving in the exclusive and nonexclusive games. Results differ between the full sample and the other two samples. The full sample results can only be explained by models that we have already rejected, while the results from the accurate responders sample or Proposition 1 compliers sample can be explained by models that we have not yet rejected. We therefore prefer the accurate responders sample and the Proposition 1 compliers sample to the full sample, since they provide a more coherent system for evaluating the role of altruism, reciprocity, and inequity aversion.

More specifically, in the full sample, giving rates are not statistically different, with P2's giving propensity equal to 40.2% in the exclusive game and 39.0% in the nonexclusive game. These results align with the model with inequity aversion (I) and the model with inequity aversion and reciprocity (IR). We have already rejected both models, since they cannot explain P1's giving behavior under Proposition 1. Using the sample of accurate responders or Proposition 1 compliers

[13] We use different subsamples in robustness checks and in almost all cases the inequality is significant.

is more promising. For the sample of accurate responders, we find significantly greater giving in the exclusive game compared with the nonexclusive game (44.1% versus 40.7%, $p < 0.05$). Similarly, the difference in P2's giving in the exclusive game is marginally significant for the sample of Proposition 1 compliers (40.8% versus 38.1%, $p < 0.10$). The results are consistent with all models that incorporate altruism. The intuition is that P2 knows that keeping in the exclusive game precludes giving by P1, and therefore definitely harms P0. However, P1 can theoretically still give in the nonexclusive game even if P2 were to keep, leading to less expected harm to P0. If P2 had altruistic concerns for P1 and P0, she should be less likely to keep in the exclusive game than in the nonexclusive game.

Finally, Propositions 7 and 8 compare P1's giving in the control group with P2's giving in the treatment groups. We find significantly greater giving by P1 in the control game than P2 in the treatment games ($p < 0.05$). In the control game, P1's giving rate is 44.7-45.5%. In the exclusive and nonexclusive games, P2's giving rates are 40.2-44.1% and 38.1-40.7%, respectively. The experimental results go against the predictions of all models that do not incorporate inequity aversion. Propositions 7 and 8 thus show that inequity aversion is necessary to align theoretical predictions with the behavior of first movers.

4.3 Summary of predictions

Table 3 summarizes our experimental results and how they align with theoretical predictions. Reciprocity incentives are necessary to establish the generalized reciprocity effect, which explains why P1's giving is higher in the treatment games exceeds than the control game (Propositions 2 and 3). Inequity aversion is necessary to explain why P2's giving rates are lower in the treatment games than P1's giving in the control game (Propositions 7 and 8). We are then left with two models: the model with inequity aversion and reciprocity (IR) and the model with altruism, inequity aversion, and reciprocity (AIR). Only the AIR model can explain why P1's giving rate in the nonexclusive game would be greater than 0 even after P2 kept (Propositions 1 and 4), making altruism a necessary component of our model as well. We conclude that altruism, inequity aversion, and reciprocity are all necessary to rationalize pay-it-forward behavior. The AIR model best rationalizes behavior for all three samples, but performs especially well in predicting behavior in the accurate responders sample, which excludes subjects who failed to demonstrate comprehension of the games, and the Proposition 1 compliers sample, which excludes subjects whose behavior was inconsistent under any model we considered.

4.4 Robustness checks

We conduct a number of robustness checks by examining results across different samples. Table 3b shows results for subjects who answered at least half of the accuracy check questions correctly on the first try. We next present t-test results for subjects who answered one or fewer questions incorrectly on the first try. The results, presented in Table 4a, show that all findings hold even though the sample reduces to only 227 subjects.

As a related robustness check, in Table 4b we restrict the sample to subjects who submitted six or fewer incorrect *answers*. The number of incorrect answers creates a different accuracy measure from the number of questions answered incorrectly, since a subject can submit multiple incorrect answers to the same question. For example, someone who submits four incorrect answers to one question but answers every other question correctly on the first try would count as having one question answered incorrectly and four incorrect answers. Again, we find that this alternative restriction does not change our main results in Table 3b.

Next, we explore whether time spent on the study affects results. We designed the study to be comfortably completed within an hour, and almost all subjects finished within 45 minutes. Excluding the 30 subjects who took more than 45 minutes from the sample of accurate responders brings the sample to 298. Despite the small sample size, our results do not change, as shown in Table 4c.

Lastly, we investigated whether the order of games affects behavior. After subjects played the control game, game order was randomized. Roughly half of the subjects were shown the nonexclusive game and then the exclusive game, and the other half played the games in the opposite order. Separately assessing the results based on order of games cuts the sample in half, yet our findings barely change. Table 4d shows results for accurate responders who saw the nonexclusive game first, while Table 4e shows results for accurate responders who saw the exclusive game first. Among the 156 subjects who saw the nonexclusive game first, empirical results are the same as in Table 3b.

Among the 168 subjects who saw the exclusive game first, all propositions still align with the AIR model predictions, but the comparisons regarding Proposition 6 become directional. Despite this, the AIR model still most consistently describes subjects' overall giving strategies across all subsamples. To see this, recall that Proposition 6 compares P2's giving rate in the exclusive and nonexclusive games. Among the accurate responders who saw the exclusive game first, we find a directional but insignificant effect. Only the inequity aversion (I) model and inequity aversion with reciprocity (IR) model would predict no significant differences in this comparison, but these models do not incorporate altruism and cannot explain other behaviors (e.g., why P1s would give after P2 keeps in the nonexclusive game). These models perform worse than the AIR model in rationalizing the totality of subjects' strategies across all nodes of our games.

4.5 Tabulating violators under each model

Another way to assess whether our experimental results align with theoretical predictions is to tabulate the proportion of individuals that violate each proposition under each model. This approach better leverages within-subject variation and generates slightly different results from Table 3 for Propositions 7 and 8, which depend on subject-level altruism and inequity aversion parameters (A, α, β). A second advantage is that unlike the paired t-tests in Table 3, this approach does not depend on analysis sample we choose, but provides an alternative method of arriving at the same conclusion.

We count violators in the following fashion. Refer to Table 1, which compares giving behavior at the two specified decision nodes under the LHS and RHS columns. When the prediction is ~ 0, the model predicts that giving rates will be statistically indistinguishable from 0, so subjects should play (keep, keep) at the two decision nodes. Anyone who plays (give, give), (give, keep), or (keep, give) would be violating the prediction of ~ 0. When the prediction is $\sim$, the model predicts equivalent actions at the two decision nodes. Anyone who plays (give, keep) or (keep, give) would violate propositions that predict $\sim$. Third, when the prediction is $\succ 0$, the model predicts that giving specified under LHS would be strictly greater than giving specified under RHS, which would be equivalent to 0. The only action that aligns with such a prediction is (give, keep), so subjects whose strategies are (give, give), (keep, keep), or (keep, give) would all be violators. Lastly, when the prediction is $\succ$, the model predicts greater giving rates at the node under LHS than under RHS. This means subjects may give at both nodes, keep at both nodes, or give at the LHS node and keep at the RHS node. The only action that violates the prediction of $\succ$ is (keep, give).

The exceptions to our method of tabulating violators are Propositions 5, 7, and 8. Since P1 plays mixed strategies, Proposition 5 predicts that P1's giving rate after P2 gives will not significantly differ between the exclusive and nonexclusive games. This does not restrict how P1's mixed strategy gets realized. Subjects who choose (give, keep) or (keep, give) are not definitive violators, since it's possible that they are indifferent between the two decisions and choose at random. Hence, (give, give), (keep, keep), (give, keep) and (keep, give) can all occur even when P1's giving rate is the same in the two games. Under any model incorporating inequity aversion, Propositions 7 and 8 depend on altruism and inequity aversion parameters (A, α, β). Depending on the specific values (A, α, β) take on, (give, give), (keep, keep), (give, keep), and (keep, give) are all plausible strategies, so all strategies for first movers are permissible if we incorporate inequity aversion into the model.

Table 5 tabulates the violators of each proposition under different models. Under model S, which assumes that subjects only care about material payoffs, as many as 64% of subjects violate Propositions 2 and 3. Models that exclude altruism (Models I and R) contain as many as 69-80%

of subjects violating at least one proposition, pointing to the necessity of altruism in explaining experimental behavior. Models excluding reciprocity also contain considerable proportions of violators. Roughly half of subjects In the model with only altruism (A), 69% of subjects in the model with only inequity aversion (I), and 30% of subjects in the model with both altruism and inequity aversion (AI) violate at least one proposition. In comparison, at most 12% of subjects in the models with both altruism and reciprocity do so.

These results (again) suggest that altruism and reciprocity are necessary components to pay-it-forward behavior. We next assess the role of inequity aversion by comparing the model with only altruism and reciprocity (AR) to the model with altruism, reciprocity, and inequity aversion (AIR). In the former, as many as 12.41% of subjects violate Proposition 8. In contrast, models that include inequity aversion do slightly better: 10.92% violate at least one proposition. Table 5 shows that using different forms of variation achieves the same conclusion as our main results in Section 4. The model that best explains our empirical results incorporates altruism, reciprocity, and inequity aversion.

4.6 Differential giving by P1 in the exclusive versus nonexclusive game: the role of perceived impact

Under all models we have considered, Proposition 5 predicts no significant differences for P1 in the exclusive and nonexclusive games. However, in the experiment we find significantly higher giving rates for P1 in the exclusive game for almost all samples (the three main samples and five supplemental samples in the robustness checks). We therefore need additional reasons to rationalize why P1's giving rates differ in the exclusive and nonexclusive games.

The exclusive game may lower P1's sense of impact on P0's payoff, since her giving decision depends on P2's prior action. P1 may believe that some of the credit for her generosity would go to P2, who made it possible for her to give in the first place. In contrast, in the nonexclusive game P1 can give to P0 independent of P2's actions, which should increase P1's credit and lower P2's credit for P0's payoff. To investigate this conjecture, we ask subjects about how much credit they attribute to each player —that is, how much of P0's payoff is due to P1's and P2's actions. Appendix Table B1 shows that subjects attribute significantly greater impact to P1 for P0's payoff in the nonexclusive game compared with the exclusive game (43.56% versus 38.74%, $p < 0.01$), and significantly greater impact to P2 for P0's payoff in the exclusive game compared to the nonexclusive game (40.77% versus 34.94%, $p < 0.01$).

If the utility function were modified to account for differential impact in the reciprocity terms, it would have the form

$$u_i(\vec{\sigma}) = \pi_i(\sigma) + A_i \sum_{j \neq i} \pi_j(\sigma) + \sum_{j \neq i} \delta_{ij} Y_i \lambda_{iji}(\vec{\sigma}) \kappa_{ij}(\vec{\sigma}) + \sum_{j,k \neq i} \delta_{ij} Z_i \lambda_{iki}(\vec{\sigma}) \kappa_{ij}(\vec{\sigma})$$
$$- \sum_{s \in S} \sigma(s) \left[\alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{\pi_j(s) - \pi_i(s), 0\} + \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{\pi_i(s) - \pi_j(s), 0\} \right],$$

and Proposition 5 would be modified in the following way.

Proposition 5'. *P1 is more/equally/less inclined to give after P2 gives in the nonexclusive game compared to the exclusive game if and only if P1's perceived impact on P0's payoff is higher/equal/lower:* $\gamma^n_{1G} \succ / \sim / \prec \gamma^e_{1G}$ *if and only if* $\delta^n_{10} > / = / < \delta^e_{10}$.

Since subjects allocate a significantly greater impact to P1 for P0's payoff in the nonexclusive game relative to the exclusive game, Proposition 5' predicts greater giving by P1 in the nonexclusive game compared to the exclusive game. Perceived impact is thus one way to reconcile our experimental results with our theoretical framework. While it may not be the only way, it is difficult to think of other reasons for P1's lower giving rates in the exclusive game compared to the nonexclusive game. For example, the variable δ^g_{10} (where $g \in \{e, n\}$) can represent how much P1 cares about P0, but P1's regard for P0 should not differ drastically between the exclusive and nonexclusive games.

5 Conclusion

Our study demonstrates that generalized reciprocity motives, altruism, and inequity aversion are all necessary to explain pay-it-forward behavior. We do this using a psychological game-theoretic framework, which formulates predictions for giving behavior under different models of fairness and altruism. We then test these predictions using a novel experiment that demonstrates the existence of generalized reciprocal exchange while controlling for alternate explanations such as income effects, relative payoffs, social image considerations. The experimental design allows us to exploit within-subject variation when comparing across various decisions in the game. That is, by assuming that subjects' utility parameters do not vary across games, we isolate distinct patterns in giving behavior in different games and player roles. Generalized reciprocity is necessary to explain why P1 gives more after P2 gives in the treatment games than in the control game ("generalized reciprocity effect"). Altruism is necessary to explain why P1 would give after P2 keeps in the exclusive game. Inequity aversion is necessary to explain why P1's giving in the control game exceeds P2's giving in the treatment games.

Our findings address the important question of how generosity spreads within communities. Generosity can spread through social networks (Fowler and Christakis, 2010), but the mechanism behind how this occurs is not well understood. We provide experimental evidence of the psychological basis for generalized reciprocal exchange, where kindness engenders further kindness. Our results speak to why people help strangers with no expectation of meeting them again, why people choose to cooperate with co-workers even at some cost to themselves, and why some communities are cooperative while others are hostile or competitive.

References

Bartlett, M. Y. and D. DeSteno (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science 17*(4), 319–325.

Battigalli, P., R. Corrao, and M. Dufwenberg (2019). Incorporating belief-dependent motivation in games. *Journal of Economic Behavior and Organization 167*, 185–218.

Battigalli, P. and M. Dufwenberg (2009). Dynamic psychological games. *Journal of Economic Theory 144*(1), 1–35.

Battigalli, P. and M. Dufwenberg (2021). Belief-dependent motivations and psychological game theory. forthcoming, Journal of Economic Literature.

Ben-Ner, A., L. Putterman, F. Kong, and D. Magan (2004). Reciprocity in a two-part dictator game. *Journal of Economic Behavior and Organization 53*(3), 333–352.

Berger, U. (2011). Learning to cooperate via indirect reciprocity. *Games and Economic Behavior 72*(1), 30–37.

Binmore, K. G. (1994). *Game theory and the social contract: playing fair*, Volume 1. MIT press.

Bolton, G. E., E. Katok, and A. Ockenfels (2005). Cooperation among strangers with limited information about reputation. *Journal of Public Economics 89*(8), 1457–1468. The Experimental Approaches to Public Economics.

Bolton, G. E. and A. Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review 90*(1), 166–193.

Charness, G., N. Du, and C.-L. Yang (2011). Trust and trustworthiness reputations in an investment game. *Games and Economic Behavior 72*(2), 361–375.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics 117*(3), 817–869.

Cox, J. C., D. Friedman, and S. Gjerstad (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior 59*(1), 17–45.

Cox, J. C., D. Friedman, and V. Sadiraj (2008). Revealed altruism. *Econometrica 76*(1), 31–69.

Desteno, D., M. Bartlett, J. Wormwood, L. Williams, and L. Dickens (2010). Gratitude as moral sentiment: Emotion-guided cooperation in economic exchange. *Emotion 10*, 289–93.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*(2), 268–298.

Ebrahimji, A. (2020, December). Over 900 cars paid for each other's meals at a Dairy Queen drive-thru in Minnesota. [Online; posted 09-December-2020].

Engelmann, D. and U. Fischbacher (2009). Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior 67*, 399–407.

Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and Economic Behavior 54*(2), 293–315.

Fehr, E. and S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives 14*(3), 159–181.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics 114*(3), 817–868.

Fowler, J. H. and N. A. Christakis (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences 107*(12), 5334–5338.

Gaudeul, A., C. Keser, and S. Müller (2021). The evolution of morals under indirect reciprocity. *Games and Economic Behavior 126*, 251–277.

Gong, B. and C.-L. Yang (2019). Cooperation through indirect reciprocity: The impact of higher-order history. *Games and Economic Behavior 118*, 316–341.

Gray, K., A. F. Ward, and M. I. Norton (2014). Paying it forward: Generalized reciprocity and the limits of generosity. *Journal of Experimental Psychology: General 143*(1), 247.

Gul, F. and W. Pesendorfer (2016). Interdependent preference models as a theory of intentions. *Journal of Economic Theory 165*, 179–208.

Heller, Y. and E. Mohlin (2017). Observations on cooperation. *Review of Economic Studies 85*(4), 2253–2282.

Herne, K., O. Lappalainen, and E. Kestilä-Kekkonen (2013). Experimental comparison of direct, general, and indirect reciprocity. *The Journal of Socio-Economics 45*, 38–46.

Hu, Y., J. Ma, Z. Luan, J. S. Dubas, and J. Xi (2019). Adolescent indirect reciprocity: Evidence from incentivized economic paradigms. *Journal of Adolescence 74*, 221–228.

Iwagami, A. and N. Masuda (2010). Upstream reciprocity in heterogeneous networks. *Journal of Theoretical Biology 265*(3), 297–305.

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

Khadjavi, M. (2017). Indirect reciprocity and charitable giving: Evidence from a field experiment. *Management Science 63*(11), 3708–3717.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics 1*(3), 593–622.

Malmendier, U., V. L. te Velde, and R. A. Weber (2014). Rethinking reciprocity. *Annual Review of Economics 6*(1), 849–874.

McCullough, M. E., M. B. Kimeldorf, and A. D. Cohen (2008). An adaptation for altruism: The social causes, social effects, and social evolution of gratitude. *Current Directions in Psychological Science 17*(4), 281–285.

Melamed, D., B. Simpson, and J. Abernathy (2020). The robustness of reciprocity: Experimental evidence that each form of reciprocity is robust to the presence of other forms of reciprocity. *Science 6*(23), eaba0504.

Mujcic, R. and A. Leibbrandt (2018). Indirect reciprocity and prosocial behaviour: Evidence from a natural field experiment. *The Economic Journal 128*(611), 1683–1699.

Nava, E., E. Croci, and C. Turati (2019). 'I see you sharing, thus I share with you': Indirect reciprocity in toddlers but not infants. *Palgrave Communications 5*(66), 1–9.

Nowak, M. A. and K. Sigmund (1998a). The dynamics of indirect reciprocity. *Journal of Theoretical Biology 194*(4), 561–574.

Nowak, M. A. and K. Sigmund (1998b). Evolution of indirect reciprocity by image scoring. *Nature 393*, 573–577.

Ohtsuki, H. and Y. Iwasa (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology 239*(4), 435–444.

Ong, D. and H. H. Lin (2011). Deserving altruism: An experiment in pure indirect reciprocity. Mimeo.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*(5), 1281–1302.

Seinen, I. and A. Schram (2006). Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review 50*(3), 581–602.

Simpson, B., A. Harrell, D. Melamed, N. Heiserman, and D. V. Negraia (2018). The roots of reciprocity: Gratitude and reputation in generalized exchange systems. *American Sociological Review 83*(1), 88–110.

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of Economic Literature 43*(2), 392–436.

Takahashi, N. (2000). The emergence of generalized exchange. *American Journal of Sociology 105*(4), 1105–1134.

Tsvetkova, M. and M. W. Macy (2014). The social contagion of generosity. *PLOS One 9*(2), e87275.

van Apeldoorn, J. and A. Schram (2016). Indirect reciprocity: A field experiment. *PLOS One 11*(4), e0152076.

Wu, J. (2018). Indirect higher order beliefs and cooperation. *Experimental Economics 21*(4), 858–876.

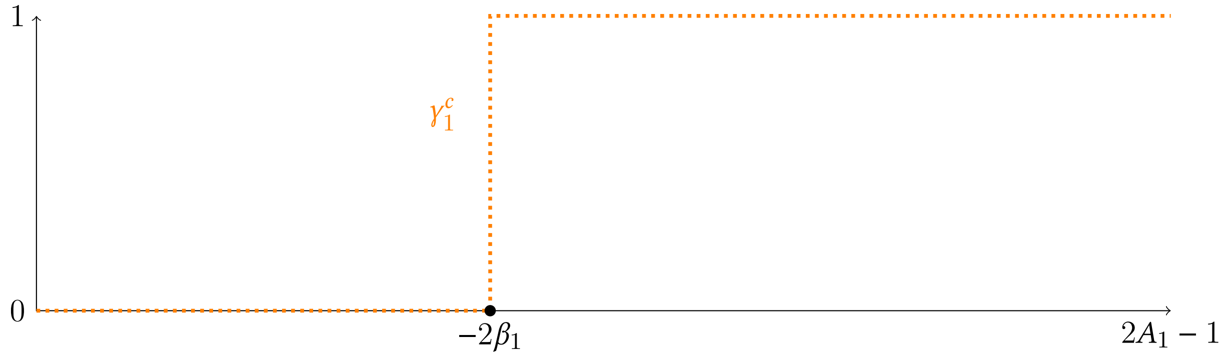
Yoeli, E., M. Hoffman, D. G. Rand, and M. A. Nowak (2013). Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences 110*(2), 10424–10429.

Zeckhauser, R., J. Swanson, and K. Lockwood (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics 9*, 79–101.

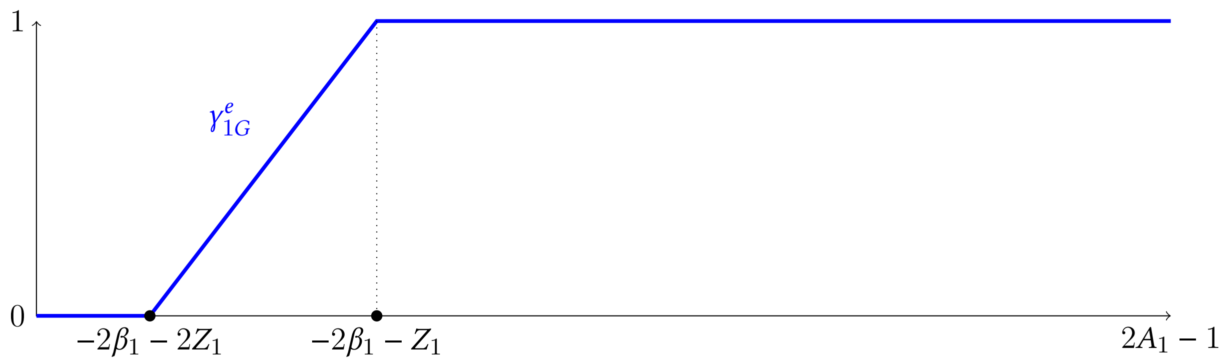
Figures & Tables

Figure 2: P1's equilibrium giving rates in different games

(a) P1's equilibrium giving rate γ_1^c in the control game



(b) P1's equilibrium giving rate γ_{1G}^e in the exclusive game



(c) P1's equilibrium giving rates γ_{1G}^n and γ_{1K}^n in the nonexclusive game

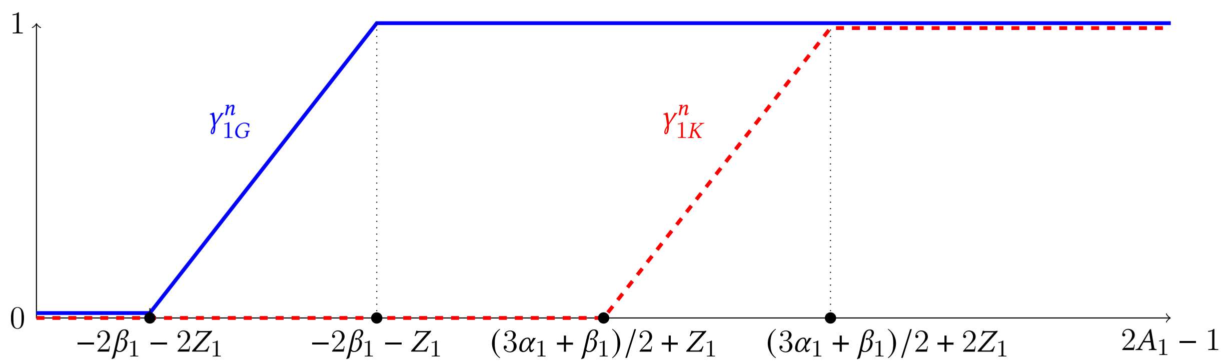


Figure 3: P1's equilibrium giving rates comparisons

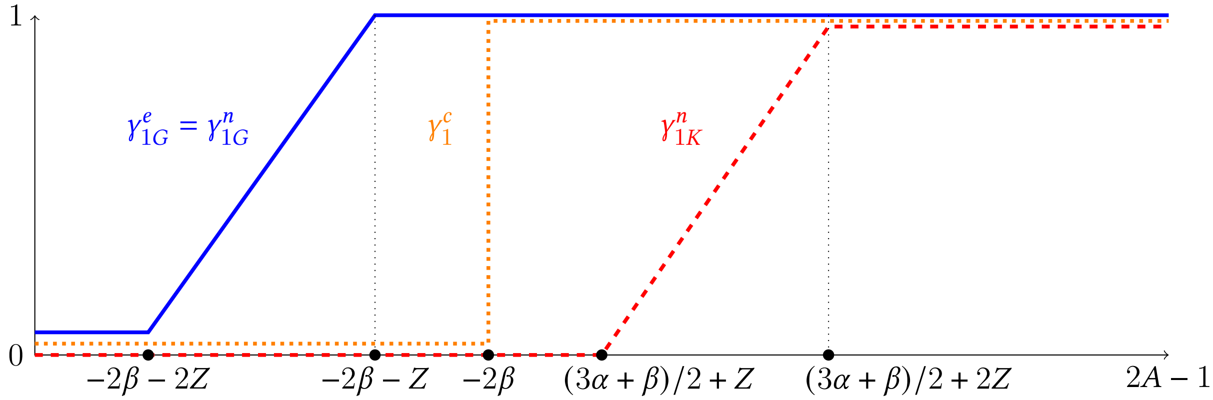


Figure 4: Initial movers' equilibrium giving rates

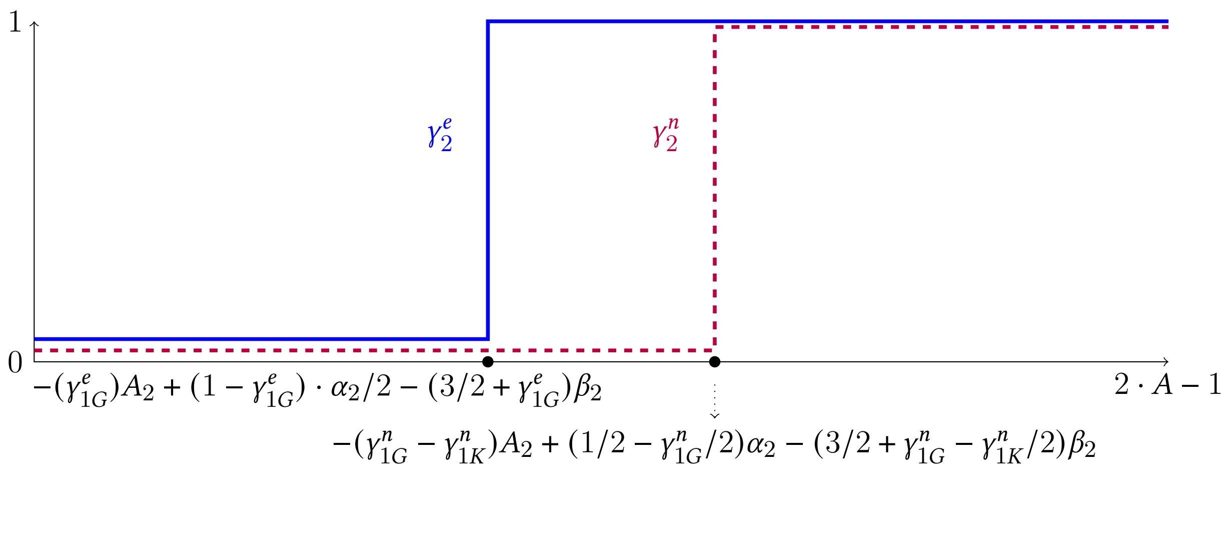


Table 1: Predicted comparisons of giving under different utilities

Model			(1) S	(2) A	(3) I	(4) R	(5) AI	(6) AR	(7) IR	(8) AIR
A_i			$= 0$	> 0	$= 0$	$= 0$	> 0	> 0	$= 0$	> 0
α_i, β_i			$= 0$	$= 0$	> 0	$= 0$	> 0	$= 0$	> 0	> 0
Z_i			$= 0$	$= 0$	$= 0$	> 0	$= 0$	> 0	> 0	> 0
Proposition	LHS	RHS	Predictions							
1	γ_{1G}^n	γ_{1K}^n	~ 0	$\sim$	> 0	> 0	$>$	$>$	> 0	$>$
2	γ_{1G}^e	γ_1^c	~ 0	$\sim$	$\sim$	> 0	$\sim$	$>$	$>$	$>$
3	γ_{1G}^n	γ_1^c	~ 0	$\sim$	$\sim$	> 0	$\sim$	$>$	$>$	$>$
4	γ_1^c	γ_{1K}^n	~ 0	$\sim$	> 0	~ 0	$>$	$>$	> 0	$>$
5	γ_{1G}^n	γ_{1G}^e	~ 0	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$	$\sim$
6	γ_2^e	γ_2^n	~ 0	$>$	$\sim$	~ 0	$>$	$>$	$\sim$	$>$
7	γ_2^e	γ_1^c	~ 0	$>$	X	~ 0	X	$>$	X	X
8	γ_2^n	γ_1^c	~ 0	$\sim$	X	~ 0	X	$>$	X	X

Note: ~ 0 indicates that LHS~RHS~0; $\sim$ indicates that LHS~RHS; > 0 indicates that LHS>RHS~0; and X indicates that the predictions are ambiguous and depend on the psychological parameters.

Table 2: Summary statistics of study subjects

	Full Sample	Accurate Responders	Prop 1 Compliers
% female	0.275	0.275	0.264
	(0.0223)	(0.0248)	(0.0231)
% college graduate	0.829	0.802	0.821
	(0.0188)	(0.0222)	(0.0201)
% employed	0.931	0.917	0.923
	(0.0127)	(0.0154)	(0.0140)
Citizenship/residency/language fluency			
% US citizen	0.684	0.673	0.675
	(0.0234)	(0.0263)	(0.0248)
% native English speaker	0.763	0.755	0.765
	(0.0213)	(0.0240)	(0.0223)
% US resident	0.727	0.701	0.717
	(0.0222)	(0.0255)	(0.0236)
Race/ethnicity			
% Black	0.129	0.0988	0.126
	(0.0167)	(0.0166)	(0.0174)
% Asian	0.293	0.306	0.294
	(0.0227)	(0.0256)	(0.0239)
% Hispanic	0.0496	0.0494	0.0495
	(0.0108)	(0.0121)	(0.0114)
% White	0.501	0.515	0.500
	(0.0249)	(0.0278)	(0.0262)
% Other race/ethnicity	0.0273	0.0309	0.0302
	(0.00813)	(0.00962)	(0.00899)
Age			
% 16-25 years old	0.159	0.160	0.154
	(0.0182)	(0.0204)	(0.0189)
% 26-35 years old	0.496	0.491	0.492
	(0.0249)	(0.0278)	(0.0262)
% 36-45 years old	0.223	0.219	0.225
	(0.0208)	(0.0230)	(0.0219)
% 46-55 years old	0.0670	0.0648	0.0714
	(0.0125)	(0.0137)	(0.0135)
% 56-65 years old	0.0422	0.0494	0.0440
	(0.0100)	(0.0121)	(0.0108)
% 65 or older	0.0124	0.0154	0.0137
	(0.00552)	(0.00686)	(0.00611)
Study characteristics			
% saw exclusive game first	0.522	0.519	0.526
	(0.0249)	(0.0278)	(0.0262)
study duration (minutes)	28.73	29.02	28.78
	(0.543)	(0.630)	(0.586)
median study duration (minutes)	26.75	26.77	26.32
bonus payment	2.553	2.731	2.580
	(0.0692)	(0.0768)	(0.0739)
median bonus payment	3	3	3
wrong answers	2.157	1.324	1.942
	(0.117)	(0.0878)	(0.117)
median wrong answers	1	1	1
Observations	403	324	364

Notes: Standard errors in parentheses.

Table 3: Experimental results of giving rates comparisons

(a) Full sample, $N = 403$

Prop	Experimental result		Consistent with predictions?							
	mean (standard error)	p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}^n_{1G} = .541(0.025)$ $> \widehat{\gamma}^n_{1K} = .233(0.021)$	$p < 0.0001$ (9.6753)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}^e_{1G} = .531(0.025)$ $> \widehat{\gamma}^c_1 = .447(0.025)$	$p = 0.0010$ (3.1112)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}^n_{1G} = .541(0.025)$ $> \widehat{\gamma}^c_1 = .447(0.025)$	$p = 0.0002$ (3.5481)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}^c_1 = .447(0.025)$ $> \widehat{\gamma}^n_{1K} = .233(0.021)$	$p < 0.0001$ (-7.1665)	No	No	No	No	Yes	Yes	No	Yes
5	$\widehat{\gamma}^n_{1G} = .541(0.025)$ $\sim \widehat{\gamma}^e_{1G} = .531(0.025)$	$p = 0.3402$ (0.4121)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	$\widehat{\gamma}^e_2 = .402(0.024)$ $\sim \widehat{\gamma}^n_2 = .390(0.024)$	$p = 0.2542$ (-0.6618)	No	D	Yes	D	D	D	Yes	D
7	$\widehat{\gamma}^e_2 = .402(0.024)$ $< \widehat{\gamma}^c_1 = .447(0.025)$	$p = 0.0156$ (-2.1612)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}^n_2 = .390(0.024)$ $< \widehat{\gamma}^c_1 = .447(0.025)$	$p = 0.0043$ (-2.6404)	No	No	Yes	No	Yes	No	Yes	Yes

Note: D indicates that the experimental results are directionally but not significantly consistent with predictions.

(b) Accurate responders, $N = 324$

Prop	Experimental result		Consistent with predictions?							
	mean (standard error)	p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}^n_{1G} = .562(0.028)$ $> \widehat{\gamma}^n_{1K} = .198(0.022)$	$p < 0.0001$ (10.7992)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}^e_{1G} = .522(0.028)$ $> \widehat{\gamma}^c_1 = .454(0.028)$	$p = 0.0057$ (2.5448)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}^n_{1G} = .562(0.028)$ $> \widehat{\gamma}^c_1 = .454(0.028)$	$p = 0.0001$ (3.9263)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}^n_{1K} = .198(0.022)$ $< \widehat{\gamma}^c_1 = .454(0.028)$	$p < 0.0001$ (-8.0592)	No	No	No	No	Yes	Yes	No	Yes
5	$\widehat{\gamma}^n_{1G} = .562(0.028)$ $> \widehat{\gamma}^e_{1G} = .522(0.028)$	$p = 0.0562$ (1.5920)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	$\widehat{\gamma}^e_2 = .441(0.028)$ $> \widehat{\gamma}^n_2 = .407(0.027)$	$p = 0.0429$ (-1.7231)	No	Yes	No	No	Yes	Yes	No	Yes
7	$\widehat{\gamma}^e_2 = .441(0.028)$ $\sim \widehat{\gamma}^c_1 = .454(0.028)$	$p = 0.2781$ (-0.5892)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}^n_2 = .407(0.027)$ $< \widehat{\gamma}^c_1 = .454(0.028)$	$p = 0.0160$ (-2.1549)	No	No	Yes	No	Yes	No	Yes	Yes

(c) Proposition 1 compliers, $N = 364$

Prop	Experimental result: mean (standard error)	Experimental result: p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}^n_{1G} = .599(0.026)$ $> \widehat{\gamma}^n_{1K} = .151(0.019)$	$p < 0.0001$ (17.157)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}^e_{1G} = .526(0.026)$ $> \widehat{\gamma}^c_1 = .455(0.026)$	$p = 0.0049$ (2.595)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}^n_{1G} = .597(0.026)$ $> \widehat{\gamma}^c_1 = .455(0.026)$	$p < 0.0001$ (5.284)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}^n_{1K} = .151(0.019)$ $< \widehat{\gamma}^c_1 = .456(0.026)$	$p < 0.0001$ (-10.914)	No	No	No	No	Yes	Yes	No	Yes
5	$\widehat{\gamma}^n_{1G} = .597(0.026)$ $> \widehat{\gamma}^e_{1G} = .526(0.026)$	$p = 0.002$ (3.10)	No	No	No	No	No	No	No	No
6	$\widehat{\gamma}^e_2 = .408(0.026)$ $> \widehat{\gamma}^n_2 = .381(0.025)$	$p = 0.0659$ (-1.51)	No	Yes	No	No	Yes	Yes	No	Yes
7	$\widehat{\gamma}^e_2 = .408(0.026)$ $< \widehat{\gamma}^c_1 = .455(0.026)$	$p = 0.0147$ (-2.188)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}^n_2 = .381(0.025)$ $< \widehat{\gamma}^c_1 = .455(0.026)$	$p = 0.0003$ (-3.51)	No	No	Yes	No	Yes	No	Yes	Yes

Columns S–AIR: Consistent with predictions?

Table 4: Experimental results of giving rates comparisons, robustness checks

(a) ≤ 1 **questions answered incorrectly** $(N = 227)$

Prop	Experimental result: mean (standard error)	Experimental result: p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}^n_{1G} = .617(0.032)$ $> \widehat{\gamma}^n_{1K} = .176(0.025)$	$p < 0.0001$ (12.1047)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}^e_{1G} = .568(0.033)$ $> \widehat{\gamma}^c_1 = .502(0.033)$	$p = 0.0055$ (2.5665)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}^n_{1G} = .617(0.032)$ $> \widehat{\gamma}^c_1 = .502(0.033)$	$p < 0.0001$ (4.1530)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}^c_1 = .502(0.033)$ $> \widehat{\gamma}^n_{1K} = .176(0.025)$	$p < 0.0001$ (-8.8330)	No	No	No	No	Yes	Yes	No	Yes
5	$\widehat{\gamma}^n_{1G} = .617(0.032)$ $> \widehat{\gamma}^e_{1G} = .568(0.033)$	$p = 0.0352$ (1.8175)	No	No	No	No	No	No	No	No
6	$\widehat{\gamma}^e_2 = .480(0.033)$ $< \widehat{\gamma}^n_2 = .511(0.033)$	$p = 0.0635$ (-1.5321)	No	Yes	No	No	Yes	Yes	No	Yes

Columns S–AIR: Consistent with predictions?

7	$\widehat{\gamma}_{2}^{e} = .511(0.033)$ $\sim \widehat{\gamma}_{1}^{c} = .502(0.033)$	$p = 0.3192$ (0.4706)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}_{2}^{n} = .480(0.033)$ $\sim \widehat{\gamma}_{1}^{c} = .502(0.033)$	$p = 0.1381$ (-1.0915)	No	Yes	Yes	No	Yes	No	Yes	Yes

Note: D indicates that the experimental results are directionally but not significantly consistent with predictions.

(b) $\leq$ 6 incorrect answers ($N = 378$)

Prop	Experimental result		Consistent with predictions?							
	mean (standard error)	p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}_{1G}^{n} = .545(0.026)$ $> \widehat{\gamma}_{1K}^{n} = .222(0.021)$	$p < 0.0001$ (10.0619)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}_{1G}^{e} = .513(0.026)$ $> \widehat{\gamma}_{1}^{c} = .450(0.026)$	$p = 0.0092$ (2.3677)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}_{1G}^{n} = .545(0.026)$ $> \widehat{\gamma}_{1}^{c} = .450(0.026)$	$p = 0.0002$ (3.5499)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}_{1K}^{n} = .222(0.021)$ $< \widehat{\gamma}_{1}^{c} = .450(0.026)$	$p < 0.0001$ (-7.4632)	No	No	No	No	Yes	Yes	No	Yes
5	$\widehat{\gamma}_{1G}^{n} = .545(0.026)$ $> \widehat{\gamma}_{1G}^{e} = .513(0.026)$	$p = 0.0954$ (1.3105)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	$\widehat{\gamma}_{2}^{e} = .423(0.025)$ $> \widehat{\gamma}_{2}^{n} = .397(0.025)$	$p = 0.0829$ (-1.3885)	No	Yes	No	No	Yes	Yes	No	Yes
7	$\widehat{\gamma}_{2}^{e} = .423(0.025)$ $\sim \widehat{\gamma}_{1}^{c} = .450(0.026)$	$p = 0.1022$ (-1.2710)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}_{2}^{n} = .397(0.025)$ $< \widehat{\gamma}_{1}^{c} = .450(0.026)$	$p = 0.0068$ (-2.4785)	No	No	Yes	No	Yes	No	Yes	Yes

(c) Accurate responders, $<$ 45 min ($N = 298$)

Prop	Experimental result		Consistent with predictions?							
	mean (standard error)	p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}_{1G}^{n} = .577(0.029)$ $> \widehat{\gamma}_{1K}^{n} = .195(0.023)$	$p < 0.0001$ (10.7353)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}_{1G}^{e} = .540(0.029)$ $> \widehat{\gamma}_{1}^{c} = .460(0.029)$	$p = 0.0020$ (2.9041)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}_{1G}^{n} = .577(0.029)$ $> \widehat{\gamma}_{1}^{c} = .460(0.029)$	$p < 0.0001$ (4.0376)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}_{1K}^{n} = .195(0.023)$ $< \widehat{\gamma}_{1}^{c} = .460(0.029)$	$p < 0.0001$ (-7.8845)	No	No	No	No	Yes	Yes	No	Yes
5	$\widehat{\gamma}_{1G}^{n} = .577(0.029)$ $> \widehat{\gamma}_{1G}^{e} = .540(0.029)$	$p = 0.0797$ (1.4107)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Prop	mean (standard error)	p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
6	$\widehat{\gamma}_2^e = .456(0.029)$ $> \widehat{\gamma}_2^n = .423(0.029)$	$p = 0.0524$ (-1.6267)	No	Yes	No	No	Yes	Yes	No	Yes
7	$\widehat{\gamma}_2^e = .456(0.029)$ $\sim \widehat{\gamma}_1^c = .460(0.029)$	$p = 0.4395$ (-0.1522)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}_2^n = .423(0.029)$ $< \widehat{\gamma}_1^c = .460(0.029)$	$p = 0.0506$ (-1.6445)	No	No	Yes	No	Yes	No	Yes	Yes

(d) Accurate responders, saw nonexclusive game first ($N = 156$)

Prop	Experimental result		Consistent with predictions?							
	mean (standard error)	p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}_{1G}^n = .551(0.040)$ $> \widehat{\gamma}_{1K}^n = .167(0.030)$	$p < 0.0001$ (7.7913)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}_{1G}^e = .487(0.040)$ $> \widehat{\gamma}_1^c = .417(0.040)$	$p = 0.0506$ (1.6488)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}_{1G}^n = .551(0.040)$ $> \widehat{\gamma}_1^c = .417(0.040)$	$p = 0.0015$ (3.0159)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}_{1K}^n = .167(0.030)$ $< \widehat{\gamma}_1^c = .417(0.040)$	$p < 0.0001$ (-5.5394)	No	No	No	No	Yes	Yes	No	Yes
5	$\widehat{\gamma}_{1G}^n = .551(0.040)$ $> \widehat{\gamma}_{1G}^e = .487(0.040)$	$p = 0.0339$ (1.8396)	No	No	No	No	No	No	No	No
6	$\widehat{\gamma}_2^e = .417(0.040)$ $> \widehat{\gamma}_2^n = .358(0.039)$	$p = 0.0302$ (-1.8921)	No	Yes	No	No	Yes	Yes	No	Yes
7	$\widehat{\gamma}_2^e = .417(0.040)$ $\sim \widehat{\gamma}_1^c = .417(0.040)$	$p = 0.5000$ (0.000)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}_2^n = .359(0.039)$ $< \widehat{\gamma}_1^c = .417(0.040)$	$p = 0.0359$ (-1.8131)	No	No	Yes	No	Yes	No	Yes	Yes

(e) Accurate responders, saw exclusive game first ($N = 168$)

Prop	Experimental result		Consistent with predictions?							
	mean (standard error)	p-value (t-stat)	S	A	I	R	AI	AR	IR	AIR
1	$\widehat{\gamma}_{1G}^n = .571(0.038)$ $> \widehat{\gamma}_{1K}^n = .226(0.032)$	$p < 0.0001$ (7.4669)	No	No	No	No	Yes	Yes	No	Yes
2	$\widehat{\gamma}_{1G}^e = .554(0.038)$ $> \widehat{\gamma}_1^c = .488(0.039)$	$p = 0.0239$ (1.9931)	No	No	No	No	No	Yes	Yes	Yes
3	$\widehat{\gamma}_{1G}^n = .571(0.038)$ $> \widehat{\gamma}_1^c = .488(0.039)$	$p = 0.0064$ (2.5137)	No	No	No	No	No	Yes	Yes	Yes
4	$\widehat{\gamma}_{1K}^n = .226(0.032)$ $< \widehat{\gamma}_1^c = .488(0.039)$	$p < 0.0001$ (-5.8375)	No	No	No	No	Yes	Yes	No	Yes

5	$\widehat{\gamma}^n_{1G} = .571(0.038)$ $\sim \widehat{\gamma}^e_{1G} = .554(0.038)$	$p = 0.3117$ (0.4921)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	$\widehat{\gamma}^e_{2} = .464(0.039)$ $\sim \widehat{\gamma}^n_{2} = .452(0.039)$	$p = 0.3194$ (-0.4703)	No	No	Yes	No	No	No	Yes	No
7	$\widehat{\gamma}^e_{2} = .464(0.039)$ $\sim \widehat{\gamma}^c_{1} = .488(0.039)$	$p = 0.1863$ (-0.8939)	No	No	Yes	No	Yes	No	Yes	Yes
8	$\widehat{\gamma}^n_{2} = .452(0.039)$ $\sim \widehat{\gamma}^c_{1} = .488(0.039)$	$p = 0.1129$ (-1.2266)	No	Yes	Yes	No	Yes	No	Yes	Yes

Table 5: Percentage of violators and violating strategy combination for each model

			(1) S	(2) A	(3) I	(4) R	(5) AI	(6) AR	(7) IR	(8) AIR
Prop	LHS	RHS	% violators and violating strategy combinations							
1	γ^n_{1G}	γ^n_{1K}	63.77% GG, GK, KG	50.12% GK, KG	59.55% GG, KK, KG	59.55% GG, KK, KG	9.68% KG	9.68% KG	59.55% GG, KK, KG	9.68% KG
2	γ^e_{1G}	γ^c_{1}	64.02% GG, GK, KG	30.27% GK, KG	30.27% GK, KG	80.65% GG, KK, KG	30.27% GK, KG	10.92% KG	10.92% KG	10.92% KG
3	γ^n_{1G}	γ^c_{1}	64.02% GG, GK, KG	29.28% GK, KG	29.28% GK, KG	80.65% GG, GK, KG	29.28% GK, KG	9.93% KG	9.93% KG	9.93% KG
4	γ^c_{1}	γ^n_{1K}	54.09% GG, GK, KG	40.20% GK, KG	69.23% GG, KK, KG	54.09% GG, GK, KG	9.43% KG	9.43% KG	69.23% GG, KK, KG	9.43% KG
5	γ^n_{1G}	γ^e_{1G}	65.27% GG, GK, KG	23.33% GK, KG	23.33% GK, KG	0% -*	23.33% GK, KG	0% -*	0% -*	0% -*
6	γ^e_{2}	γ^n_{2}	46.65% GG, GK, KG	6.45% KG	14.14% GK, KG	46.65% GG, GK, KG	6.45% KG	6.45% KG	14.14% GK, KG	6.45% KG
7	γ^e_{2}	γ^c_{1}	51.12% GG, GK, KG	10.92% KG	6.45% GK**	51.12% GG, GK, KG	6.45% GK**	10.92% GK**	6.45% GK**	6.45% GK**
8	γ^n_{2}	γ^c_{1}	51.36% GG, GK, KG	12.41% KG	6.70% GK**	51.36% GG, GK, KG	6.70% GK**	12.41% KG	6.70% GK**	6.70% GK**

*: Players can play mixed strategies, so all strategy combinations can comply with the predictions. Players playing GK or KG are potential violators.
**: Violators play either GK or KG, depending on the parameters of the model. We report the lower percentage of violators in the table.

A Omitted proofs

A.1 Remark on efficient strategies

For technical reasons, when we define kindness we ignore Pareto-inefficient strategies and focus on Pareto-efficient ones. In our experiments, players do not have inefficient strategies. For the sake of completeness for the theory and to be consistent with previous theories, we keep this assumption. Intuitively, a strategy is inefficient if another strategy provides (i) no lower material payoff for any player for any history of play and the subsequent choices of others and (ii) a strictly higher payoff for some player for some history of play and subsequent choices by the others. Formally, player i's set of efficient strategies is

$$\Sigma_i^e := \Bigg\{ \sigma_i \in \Sigma_i \Bigg| \nexists \widehat{\sigma}_i \in \Sigma_i \text{ such that } \forall h \in H, \sigma_{-i} \in \Sigma_{-i}, k \in N,$$

$$\pi_k(\widehat{\sigma}_i(h), \sigma_{-i}(h)) \geq \pi_k(\sigma_i(h), \sigma_{-i}(h)) \quad \text{with strict inequality for some } (h, \sigma_{-i}, k) \Bigg\}.$$

A.2 Proof of the theorem on equilibrium existence

Proof of Theorem 1. Let $\Sigma_i(h)$ denote i's set of (potentially random) choices at history $h \in H$. For any $s \in \Sigma_i(h)$, let $\sigma_i(h, s)$ denote player i's strategy that specifies the choice s at h, but is the same as $\sigma_i(h)$ otherwise—i.e., at every history in $H \backslash \{h\}$. Define correspondence $\beta_{i,h} : \Sigma \to \Sigma_i(h)$ by

$$\beta_{i,h}(\sigma) = \arg\max_{x \in X_i(h)} u_i(\sigma_i(h, s), (\sigma_j(h), (\sigma_k(h))_{k \neq j})_{j \neq i}),$$

and define correspondence $\beta : \Sigma \to \Pi_{(i,h) \in N \times H} \Sigma_i(h)$ by

$$\beta(\sigma) = \prod_{(i,h) \in N \times H} \beta_{i,h}(\sigma).$$

The set $\prod_{(i,h) \in N \times H} \Sigma_i(h)$ is topologically equivalent to the set Σ, so $\beta : \Sigma \to \Pi_{(i,h) \in N \times H} \Sigma_i(h)$ is equivalent to a correspondence $\gamma : \Sigma \to \Sigma$ (which is a direct redefinition of β). Every fixed point of γ is an equilibrium. To see this, note that a fixed point $\beta_{i,h}$ satisfies utility maximization under consistent beliefs. Here, because $\beta_{i,h}$ specifies the optimal choices at each $h \in H$, altogether, $\beta_{i,h}$ specifies the optimal strategies in $\Sigma_i(h, s)$. Hence, β and γ are combined best-response correspondences. Since γ is a correspondence from Σ to Σ, it is amenable to fixed-point analysis.

It remains to show that γ possesses a fixed point. Berge's maximum principle guarantees that $\beta_{i,h}$ is nonempty, closed-valued, and upper hemicontinuous, since $\Sigma_i(h)$ is nonempty and compact and u_i is continuous (since π_i, κ_{ij}, and λ_{ijk} are all continuous). In addition, $\beta_{i,h}$ is convex-valued, since $\Sigma_i(h)$ is convex and u_i is linear—and hence quasiconcave—in i's own choice. Hence, $\beta_{i,h}$ is nonempty, closed-valued, upper hemicontinuous, and convex-valued. These properties extend to β and γ. Hence, it follows by Kakutani's fixed-point theorem that γ admits a fixed point. □

A.3 Proof of lemmas on equilibrium giving strategy

Proof of Lemma 1. P1's choice is between giving, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 2, 2)$, and keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$. P1's utility from giving is $u_1(g_1) = 2+4\cdot A_1$. P1's utility from keeping is $u_1(k_1) = 3+A_1\cdot 2-\beta_1\cdot(3-2)/2-\beta_1\cdot(3-0)/2 = 3+2\cdot A_1-2\cdot\beta_1$. P1 prefers giving if and only if $u_1(g_1) = 2+4\cdot A_1 \geq u_1(k_1) = 3+2\cdot A_1-2\cdot\beta_1$, that is, $2\cdot A_1+2\cdot\beta_1 \geq 1$. □

Proof of Lemma 2. Suppose P1 believes that P2 believes that P1 gives with probability γ''_{1G}. The equitable payoff of P1 is $(1+3-\gamma''_{1G})/2 = 2-\gamma''_{1G}/2$, so giving by P2 to P1 shows a kindness of $3-\gamma''_{1G}-(2-\gamma''_{1G}/2) = 1-\gamma''_{1G}/2$. P1's utility from giving, which results in material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 2, 2)$, is $u_1(g_{1G}, \gamma''_{1G}) = 2+4\cdot A_1+Z_1\cdot(+1)\cdot(1-\gamma''_{1G}/2)$, and P1's utility from keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$, is $u_1(k_{1G}, \gamma''_{1G}) = 3+2\cdot A_1-2\cdot\beta_1+Z_1\cdot(-1)\cdot(1-\gamma''_{1G}/2)$. Therefore, P1's utility from giving with probability γ_{1G} is $u_1(\gamma_{1G}, \gamma''_{1G}) = \gamma_{1G}\cdot[-1+2\cdot A_1+2\cdot\beta_1+Z_1\cdot(2-\gamma''_{1G})]+3+2\cdot A_1-2\cdot\beta_1-Z_1\cdot(1-\gamma''_{1G}/2)$. If $2\cdot A_1+2\cdot\beta_1+Z_1 \geq 1$, then $\gamma_{1G} = 1$. If $2\cdot A_1+2\cdot\beta_1+2\cdot Z_1 \leq 1$, then $\gamma_{1G} = 0$. If $2\cdot A_1+2\cdot\beta_1+Z_1 < 1 < 2\cdot A_1+2\cdot\beta_1+2\cdot Z_1$, then $-1+2\cdot A_1+2\cdot\beta_1+Z_1\cdot(2-\gamma_{1G}) = 0$, which rearranges to $\gamma_{1G} = 2-(1-2\cdot A_1-2\cdot\beta_1)/Z_1$. Therefore, in equilibrium, $\gamma^e_{1G} = [\![2+(2\cdot A_1+2\cdot\beta_1-1)/Z_1]\!]$.

Suppose P2 believes that P1 gives with probability γ'_{1G}. P2's expected utility from giving, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 2, 2)$ with probability γ'_{1G} and $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$ with probability $1-\gamma'_{1G}$, is $\gamma'_{1G}\cdot(2+4\cdot A_2)+(1-\gamma'_{1G})\cdot(2+3\cdot A_2-\alpha_2/2-\beta_2) = 2+4\cdot A_2-(1-\gamma'_{1G})\cdot(A_2+\alpha_2/2+\beta_2)$. P2's utility from keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$, is $3+1\cdot A_2-\beta_2\cdot(3-1)/2-\beta_2\cdot(3-0)/2 = 3+A_2-5\cdot\beta_2/2$. P2 prefers giving if $2+4\cdot A_2-(1-\gamma'_{1G})\cdot(A_2+\alpha_2/2+\beta_2) \geq 3+A_2-5\cdot\beta_2/2$, which is simplified to $3\cdot A_2+5\cdot\beta_2/2-(1-\gamma'_{1G})\cdot(A_2+\alpha_2/2+\beta_2) \geq 1$. In equilibrium, $\gamma'_{1G} = \gamma_{1G}$, so the inequality is rearranged to $(2+\gamma_{1G})\cdot A_2-(1-\gamma_{1G})\cdot\alpha_2/2+(3/2+\gamma_{1G})\cdot\beta_2 \geq 1$. □

Proof of Lemma 3. Suppose P1 believes that P2 believes that P1 gives with probability γ''_{1G} when P2 gives, and gives with probability γ''_{1K} when P2 keeps. First, suppose P2 keeps. P0's equitable payoff is 1, and P1's equitable payoff is $[(1-\gamma''_{1K})+(3-\gamma''_{1G})]/2 = 2-\gamma''_{1K}/2-\gamma''_{1G}/2$.

First, consider when P2 keeps. P1's utility from giving, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (3, 0, 2)$, is $u_1(k_2, g_{1K}, \cdots) = 0+5\cdot A_1-\alpha_1\cdot(3-0)/2-\alpha_1\cdot(2-0)/2+Z_1\cdot(+1)\cdot(\gamma''_{1G}/2-1-\gamma''_{1K}/2) = 5\cdot A_1-5\cdot\alpha_1/2+Z_1\cdot(\gamma''_{1G}/2-1-\gamma''_{1K}/2)$, and P1's utility from keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$, is $u_1(k_2, k_{1K}, \cdots) = 1+3\cdot A_1-\alpha_1\cdot(3-1)/2-\beta_1\cdot(1-0)/2-$

$Z_1 \cdot (\gamma''_{1G}/2-1-\gamma''_{1K}/2) = 1+3\cdot A_1-3\cdot\alpha_1/2-\beta_1/2-Z_1\cdot(\gamma''_{1G}/2-1-\gamma''_{1K}/2)$. Fixing γ''_{1G} and γ''_{1K}, we have $u_1(k_2, g_{1K}, \cdots) - u_1(k_2, k_{1K}, \cdots) = -1+2\cdot A_1-3\cdot\alpha_1/2-\beta_1/2+2\cdot Z_1\cdot(\gamma''_{1G}/2-1-\gamma''_{1K}/2) = -1+2\cdot A_1-3\cdot\alpha_1/2-\beta_1/2-Z_1\cdot(2-\gamma''_{1G}+\gamma''_{1K})$.

Second, consider when P2 gives. Regarding the reciprocity payoff, the only change is in the flip of the sign of λ_{121}. P1's utility of giving, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 2, 2)$, is $u_1(g_2, g_{1G}, \cdots) = 2+4\cdot A_1+Z_1\cdot(1-\gamma''_{1G}/2+\gamma''_{1K}/2)$. P1's utility of keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$, is $u_1(g_2, k_{1G}, \cdots) = 3+2\cdot A_1-\beta_1\cdot(3-2)/2-\beta_1\cdot(3-0)/2-Z_1\cdot(1-\gamma''_{1G}/2+\gamma''_{1K}/2) = 3+2\cdot A_1-2\cdot\beta_1-Z_1\cdot(1-\gamma''_{1G}/2+\gamma''_{1K}/2)$. Hence, $u_1(g_2, g_{1G}, \cdots) - u_1(g_2, k_{1G}, \cdots) = -1+2\cdot A_1+2\cdot\beta_1+Z_1\cdot(2-\gamma''_{1G}+\gamma''_{1K})$.

Comparing the net benefit of giving after P2 gives and that after P2 keeps, we have

$$u_1(g_2, g_{1G}, \cdots) - u_1(g_2, k_{1G}, \cdots) \geq u_1(k_2, g_{1G}, \cdots) - u_1(k_2, k_{1G}, \cdots).$$

Hence, whenever P1 decides to give after P2 keeps, she will also choose to give after P2 gives. In other words, P1 is more inclined to give after P2 gives than after P2 keeps: $\gamma_{1G} \geq \gamma_{1K}$. Given this inequality, there are five possible cases regarding γ_{1G} and γ_{1K}.

1. Strategies $\gamma_{1G} = 1$ and $\gamma_{1K} = 1$ are supported in equilibrium when and only when $2\cdot A_1 - 3\cdot\alpha_1/2-\beta_1/2-2\cdot Z_1 \geq 1$.

2. Strategies $\gamma_{1G} = 1$ and $0 < \gamma_{1K} < 1$ are supported in equilibrium when and only when $2\cdot A_1-3\cdot\alpha_1/2-\beta_1/2-2\cdot Z_1 \leq 1 \leq 2\cdot A_1-3\cdot\alpha_1/2-\beta_1/2-Z_1$. In this case, $2\cdot A_1-3\cdot\alpha_1/2-\beta_1/2-Z_1\cdot(1+\gamma_{1K}) = 1$, which is rearranged to $\gamma_{1K} = (2\cdot A_1-3\cdot\alpha_1/2-\beta_1/2-1-Z_1)/Z_1$.

3. Strategies $\gamma_{1G} = 1$ and $\gamma_{1K} = 0$ are supported in equilibrium when and only when $2\cdot A_1 - 3\cdot\alpha_1/2-\beta_1/2-Z_1 \leq 1 \leq 2\cdot A_1+2\cdot\beta_1+Z_1$.

4. Strategies $0 < \gamma_{1G} < 1$ and $\gamma_{1K} = 0$ are supported in equilibrium when and only when $2\cdot A_1+2\cdot\beta_1+Z_1 \leq 1 \leq 2\cdot A_1+2\cdot\beta_1+2\cdot Z_1$. In this case, $2\cdot A_1+2\cdot\beta_1+Z_1\cdot(2-\gamma_{1G}) = 1$, which is rearranged to $\gamma_{1G} = 2+(2\cdot A_1+2\cdot\beta_1-1)/Z_1$.

5. Strategies $\gamma_{1G} = 0$ and $\gamma_{1K} = 0$ are supported in equilibrium when and only when $2\cdot A_1 + 2\cdot\beta_1+2\cdot Z_1 \leq 1$.

In summary, in the nonexclusive game, P1 gives with probability $\gamma^n_{1G} = [\![(2\cdot A_1+2\cdot\beta_1-1)/Z_1+2]\!]$ after P2 gives, and gives with probability $\gamma^n_{1K} = [\![(2\cdot A_1-\alpha_1-\beta_1/2-1)/Z_1-1]\!]$ after P2 keeps.

Consider P2's action next. Suppose P2 believes that P1 gives with probability γ'_{1G} and γ'_{1K} when P2 gives and keeps, respectively. P2's expected utility from giving, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 2, 2)$ with probability γ'_{1G} and $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$ with probability

$1-\gamma'_{1G}$, is $(2+4\cdot A_2)\cdot\gamma'_{1G}+(2+3\cdot A_2-\alpha_2/2-\beta_2)\cdot(1-\gamma'_{1G}) = 2+(3+\gamma'_{1G})\cdot A_2-(1-\gamma'_{1G})\cdot(\alpha_2/2+\beta_2)$. P2's expected utility from keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (3, 0, 2)$ with probability γ'_{1K} and $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$ with probability $1-\gamma'_{1K}$, is $\gamma'_{1K}\cdot[3+2\cdot A_2-\beta_2\cdot(3-0)/2-\beta_2\cdot(3-2)/2]+(1-\gamma'_{1K})\cdot[3+A_2-\beta_2\cdot(3-0)/2-\beta_2\cdot(3-1)/2] = 3+(1+\gamma'_{1K})\cdot A_2-(5-\gamma'_{1K})\cdot\beta_2/2$. P2 prefers giving if and only if $2+(3+\gamma'_{1G})\cdot A_2-(1-\gamma'_{1G})\cdot(\alpha_2/2+\beta_2) \geq 3+(1+\gamma'_{1K})\cdot A_2-(5-\gamma'_{1K})\cdot\beta_2/2$, which, as $\gamma_{1G} = \gamma'_{1G}$ and $\gamma_{1K} = \gamma'_{1K}$ in equilibrium, is rearranged to $(2 + \gamma_{1G} - \gamma_{1K}) \cdot A_2 - (1/2 - \gamma_{1G}/2) \cdot \alpha_2 + (3/2 + \gamma_{1G} - \gamma_{1K}/2) \cdot \beta_2 \geq 1$. □

A.4 Proofs of propositions on equilibrium giving comparisons

Proof of Proposition 1. In the nonexclusive game, P1 gives with probability $\gamma^n_{1G} = [\![(2\cdot A_1+2\cdot\beta_1-1)/Z_1+2]\!]$ after P2 gives, and P1 gives with probability $\gamma^n_{1K} = [\![(2\cdot A_1-3\cdot\alpha_1/2-\beta_1/2-1)/Z_1-1]\!]$ after P2 keeps. Since $(2 \cdot A_1 + 2 \cdot \beta_1 - 1)/Z_1 + 2 > (2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - 1)/Z_1 - 1$ for any combination of nonnegative parameters A_1, α_1, β_1, and Z_1, $\gamma^n_{1G} \geq \gamma^n_{1K}$, and the inequality is strict as long as $Z_1 \neq 0$. □

Proof of Proposition 2. Explicitly, P1's equilibrium probability of giving in the control game is

$$\gamma^c_1 = \begin{cases} 1 & \text{if } 2 \cdot A_1 + 2 \cdot \beta_1 > 1, \\ 0 & \text{if } 2 \cdot A_1 + 2 \cdot \beta_1 < 1, \end{cases}$$

and P1's equilibrium probability of giving in the exclusive game is

$$\gamma^e_{1G} = \begin{cases} 1 & \text{if } 2 \cdot A_1 + 2 \cdot \beta_1 + Z_1 > 1, \\ (2 \cdot A_1 + 2 \cdot \beta_1 - 1)/Z_1 + 2 & \text{if } 2 \cdot A_1 + 2 \cdot \beta_1 + 2 \cdot Z_1 \geq 1 \geq 2 \cdot A_1 + 2 \cdot \beta_1 + Z_1, \\ 0 & \text{if } 2 \cdot A_1 + 2 \cdot \beta_1 + 2 \cdot Z_1 < 1. \end{cases}$$

When $Z_1 = 0$, the condition for $\gamma^e_{1G} = 1$ and the condition for $\gamma^c_1 = 1$ coincide, and the condition for $\gamma^e_{1G} = 0$ and the condition for $\gamma^c_1 = 0$ also coincide. When $Z_1 > 0$, the set of parameters for $\gamma^e_{1G} = 1$ is a strict superset of that for $\gamma^c_1 = 1$, and the set of parameters for $\gamma^e_{1G} = 0$ is a strict subset of that for $\gamma^c_1 = 0$. For the set range of parameters for $0 < \gamma^e_{1G} < 1$, $\gamma^c_1 = 0$. Hence, $\gamma^e_{1G} \geq \gamma^c_1$ for any combination of parameters. Hence, $\gamma^e_{1G} \succ \gamma^c_1$ □

Proof of Proposition 3. The proof mimics the proof of Proposition 2, with superscripts e replaced by superscripts n. Alternatively, by Proposition 5, $\gamma^e_{1G} \sim \gamma^n_{1G}$, so by transitivity of the inclination, $\gamma^n_{1G} \succ \gamma^c_1$. □

Proof of Proposition 4. Explicitly, P1's equilibrium probability of giving in the control game is

$$\gamma_1^c = \begin{cases} 1 & \text{if } 2 \cdot A_1 + 2 \cdot \beta_1 > 1, \\ 0 & \text{if } 2 \cdot A_1 + 2 \cdot \beta_1 < 1, \end{cases}$$

and P1's equilibrium probability of giving after P2 keeps in the nonexclusive game is

$$\gamma_{1K}^n = \begin{cases} 1 & \text{if } 2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - 2 \cdot Z_1 > 1, \\ (2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - 1)/Z_1 - 1 & \text{if } 2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - 2 \cdot Z_1 \leq 1 \\ & \quad \leq 2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - Z_1 \\ 0 & \text{if } 2 \cdot A_1 - 3 \cdot \alpha_1/2 - \beta_1/2 - Z_1 < 1. \end{cases}$$

When $\alpha_1 = \beta_1 = Z_1 = 0$, the two decisions coincide. When $\alpha_1 > 0$, $\beta_1 > 0$, and/or $Z_1 > 0$, the set of parameters for $\gamma_1^c = 1$ is a strict superset of that for $\gamma_1^{1K} = 1$, and the set of parameters for $\gamma_1^c = 0$ is a strict subset of that for $\gamma_1^{1K} = 0$. Hence, $\gamma_1^c > \gamma_{1K}^n$. □

Proof of Proposition 5. P1 gives with probability $\gamma_{1G}^e = [\![(2 \cdot A_1 + 2 \cdot \beta_1 - 1)/Z_1 + 2]\!]$ after P2 gives in the exclusive game. Equally, P1 gives with probability $\gamma_{1G}^e = [\![(2 \cdot A_1 + 2 \cdot \beta_1 - 1)/Z_1 + 2]\!]$ after P2 gives in the nonexclusive game. Hence, P1 is equally inclined to give in the two treatment games after P2 gives. □

Proof of Proposition 6. As shown by the inequality condition in Lemma 2 that characterizes P2's preference for giving, P2's net benefit of giving over keeping in the exclusive game is $B^e \equiv (2 + \gamma_{1G}^e) \cdot A_2 - (1/2 - \gamma_{1G}^e/2) \cdot \alpha_2 + (3/2 + \gamma_{1G}^e) \cdot \beta_2 - 1$. Similarly, by the inequality condition in Lemma 3 that characterizes P2's preference for giving, P2's net benefit of giving over keeping in the nonexclusive game is $B^n \equiv (2 + \gamma_{1G}^n - \gamma_{1K}^n) \cdot A_2 - (1/2 - \gamma_{1G}^n/2) \cdot \alpha_2 + (3/2 + \gamma_{1G}^n - \gamma_{1K}^n) \cdot \beta_2 - 1$. By Proposition 5, $\gamma_{1K}^n = \gamma_{1G}^n$. Then, $B^e - B^n = \gamma_{1K}^n \cdot A_2 + \gamma_{1K}^n \cdot \beta_2$. Since $A_1 \geq 0$ and $\beta_1 \geq 0$ in the general AIR utility function, and $\gamma_{1K}^n > 0$ in equilibrium, $B^e - B^n \geq 0$. The higher net benefit of giving over keeping in the exclusive game implies a higher inclination of giving in the exclusive game than the nonexclusive game. □

Proof of Proposition 7. By the inequality condition in Lemma 1 that characterizes P1's preference for giving, P1's net benefit of giving over keeping in the control game is $B^c \equiv 2 \cdot A_1 + 2 \cdot \beta_1 - 1$. As shown by the inequality condition in Lemma 2 that characterizes P2's preference for giving, P2's net benefit of giving over keeping in the exclusive game is $B^e \equiv (2 + \gamma_{1G}^e) \cdot A_2 - (1/2 - \gamma_{1G}^e/2) \cdot \alpha_2 + (3/2 + \gamma_{1G}^e) \cdot \beta_2 - 1$. For the same subject, i.e., $A_1 = A_2 \equiv A$, $\alpha_1 = \alpha_2 \equiv \alpha$, $\beta_1 = \beta_2 \equiv \beta$, $Z_1 = Z_2 \equiv Z$, the difference in the net benefits is $B^e - B^c = \gamma_{1G}^e \cdot A - (1 - \gamma_{1G}^e) \cdot \alpha/2 + (\gamma_{1G}^e - 1/2) \cdot \beta$.

Therefore, P1 in the control game is more inclined to give than P2 in the exclusive game, if and only if $B^e - B^c \leq 0$, that is, $\gamma^e_{1G} \cdot A + (\gamma^e_{1G} - 1/2) \cdot \beta \leq (1 - \gamma^e_{1G}) \cdot \alpha/2$. □

Proof of Proposition 8. By the inequality condition in Lemma 1 that characterizes P1's preference for giving, P1's net benefit of giving over keeping in the control game is $B^c \equiv 2 \cdot A_1 + 2 \cdot \beta_1 - 1$. By the inequality condition in Lemma 3 that characterizes P2's preference for giving, P2's net benefit of giving over keeping in the nonexclusive game is $B^n \equiv (2 + \gamma^n_{1G} - \gamma^n_{1K}) \cdot A_2 - (1/2 - \gamma^n_{1G}/2) \cdot \alpha_2 + (3/2 + \gamma^n_{1G} - \gamma^n_{1K}) \cdot \beta_2 - 1$. For the same subject, i.e., $A_1 = A_2 \equiv A$, $\alpha_1 = \alpha_2 \equiv \alpha$, $\beta_1 = \beta_2 \equiv \beta$, $Z_1 = Z_2 \equiv Z$, the difference in the net benefits is $B^n - B^c = (\gamma^n_{1G} - \gamma^n_{1K}) \cdot A - (1/2 - \gamma^n_{1G}/2) \cdot \alpha + (\gamma^n_{1G} - \gamma^n_{1K} - 1/2) \cdot \beta$. Therefore, P1 in the control game is more inclined to give than P2 in the nonexclusive game, if and only if $B^n - B^c \leq 0$, that is, $(\gamma^n_{1G} - \gamma^n_{1K}) \cdot A + (\gamma^n_{1G} - \gamma^n_{1K} - 1/2) \cdot \beta \leq (1/2 - \gamma^n_{1G}/2) \cdot \alpha$. □

Proof of Proposition 5'. Modifying the proofs of Lemmas 2 and 3, the equilibrium giving rate when agents' preferences incorporate credit-based reciprocity is $\gamma^g_{1G} = [\![(2 \cdot A_1 + 2 \cdot \beta_1 - 1)/(\delta^g_{10} \cdot Z_1) + 2]\!]$ for game $g \in \{e, n\}$. When the giving rate is between 0 and 1, $(2 \cdot A_1 + 2 \cdot \beta_1 - 1)/(\delta^g_{10} \cdot Z_1)$ is between -2 and -1. Hence, in this range, as δ^g_{10} increases, the term increases. Therefore, $\gamma^n_{1G} \geq \gamma^e_{1G}$ if and only if $\delta^n_{10} \geq \delta^e_{10}$. □

B Additional experimental results

Table B1: Credit for each player in each game

	Control	Exclusive	Nonexclusive
P2's Credit for P2's Payoff		59.45	59.29
		(1.455)	(1.478)
P2's Credit for P1's Payoff	22.01	44.40	43.33
	(0.892)	(0.947)	(0.868)
P2's Credit for P0's Payoff	22.33	40.77	34.94
	(0.955)	(0.926)	(0.934)
P1's Credit for P2's Payoff		21.93	22.23
		(0.849)	(0.896)
P1's Credit for P1's Payoff	56.74	36.91	36.62
	(1.448)	(0.779)	(0.800)
P1's Credit for P0's Payoff	54.39	37.62	42.53
	(1.501)	(0.944)	(1.206)
P0's Credit for P2's Payoff		18.62	18.49
		(0.866)	(0.868)
P0's Credit for P1's Payoff	21.25	18.69	20.05
	(0.919)	(0.874)	(0.927)
P0's Credit for P0's Payoff	23.28	21.61	22.53
	(1.047)	(1.064)	(1.077)

Standard errors in parentheses.