Pay it forward: Theory and experiment ☆

Amanda Chuan [a],[ID],[*], Hanzhe Zhang [b],[ID],[*]

[a] School of Human Resources and Labor Relations, Michigan State University, United States of America
[b] Department of Economics, Michigan State University, United States of America

ARTICLE INFO

ABSTRACT

We theoretically and experimentally investigate psychological motivations behind pay-it-forward behavior. We construct a psychological game-theoretic model that incorporates altruism, inequity aversion, and indirect reciprocity following Rabin (1993), Fehr and Schmidt (1999), and Dufwenberg and Kirchsteiger (2004). We test this model using games in which players choose to give to strangers, potentially after receiving a gift from an unrelated benefactor. Our experiment reveals that altruism and indirect reciprocity spur people to pay kind actions forward, informing how kindness begets further kindness. However, inequity aversion hinders giving even when giving will allow one's kindness to be paid forward. Our paper informs how kind behaviors get passed on among parties that never directly interact, which has implications for the formation of social norms and behavioral conduct within workplaces, neighborhoods, and communities.
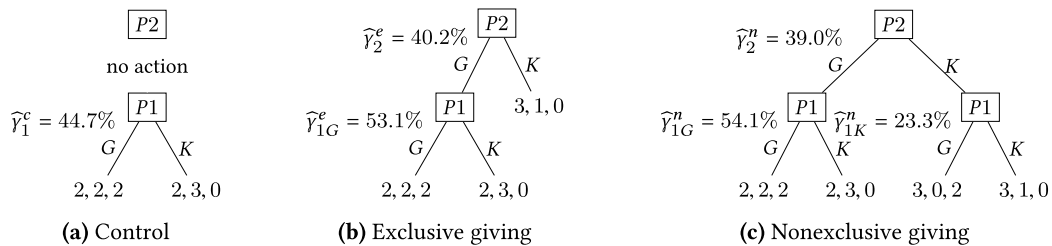
1. Introduction

Pay-it-forward behavior is at the heart of various social exchanges, ranging from the everyday to the life-saving. An employee who was mentored by a superior may elect to advise a new coworker in turn. In fast food drive-throughs, when a customer learns that the previous customer paid for her meal, she is more likely to pay for the meal of the customer after her in line. One such transaction in a Minnesota Dairy Queen culminated in a chain of giving that lasted 900 cars long (Ebrahimji, 2020). In organ exchange, individuals donate their kidneys to strangers if their loved ones receive a kidney from a compatible donor, creating exchange chains that save hundreds of lives (Roth et al., 2004).

How do we start and maintain chains of giving? Maintaining these chains may be natural and automatic if receiving a gift makes you more likely to give to an unrelated third party. These chains will readily start if the knowledge that others may pay your gift forward, expanding your impact, increases the likelihood that you give. To test these hypotheses, we run a laboratory experiment and guide our observations with a psychological game-theoretic model. While prior work has established evidence for pay-it-forward behavior in lab and field settings (Ben-Ner et al., 2004; Bartlett and DeSteno, 2006; Desteno et al., 2010; Herne et al., 2013; Gray et

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

Note: Each player's index denotes the number of players behind them in the giving chain: P0 is the last potential recipient, P1 is the last player to decide on giving, and P2, if allowed, is the penultimate player to decide on giving. Material payoffs are (π_2, π_1, π_0). The $\hat{\gamma}$s indicate the giving rates (probabilities of choosing G) in our experiment. The superscript denotes game type, where c stands for control, e for exclusive, and n for nonexclusive. The subscript 1 denotes P1's action, and the subscript 2 denotes P2's action. The subscript G stands for P1's decision after P2 gave, and K for P1's decision after P2 kept.

Fig. 1. Games and giving rates in the experiment.

al., 2014; Tsvetkova and Macy, 2014; van Apeldoorn and Schram, 2016; Mujcic and Leibbrandt, 2018; Simpson et al., 2018; Melamed et al., 2020), to our knowledge, we are the first to investigate the psychological motivations that support pay-it-forward behavior. Addressing the psychological underpinnings of pay-it-forward behavior is important for understanding key levers that promote the transmission of kindness.

Fig. 1 depicts the three three-player games in the experiment. In all three games, players choose whether to pass a chip worth $1 to the next player (give, G), or not (keep, K). Following the multiplier methods used for investment and public goods games, a passed chip turns into two chips.[1]

Fig. 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 Figs. 1b and 1c, P2 is endowed with 3 chips, P1 with 1 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 same 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.

In the treatment games, P2's decisions impact P1 directly and P0 indirectly. We vary the extent of P2's indirect impact on P0 in the *exclusive giving* and *nonexclusive giving* conditions. In the *exclusive* game depicted by Fig. 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 Fig. 1c, P1 can give P0 a chip regardless of whether P2 gives a chip to P1 first. If P1 chooses to keep after P2 kept, the game concludes with payoffs $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$. If P1 gives even after P2 kept, the game concludes with payoffs $(\pi_2, \pi_1, \pi_0) = (3, 0, 2)$.

We start by comparing the giving decisions of Last Movers, P1 in all games, to establish evidence of pay-it-forward behavior: Receiving a gift makes a subject more likely to give to a third party. As shown by the empirical giving rates $\hat{\gamma}$ in Fig. 1, P1 is most likely to give when P2 has given in the nonexclusive game ($\hat{\gamma}_{1G}^n = 54.1\%$) or in the exclusive game ($\hat{\gamma}_{1G}^e = 53.1\%$); less likely to give when P2 cannot give in the control game ($\hat{\gamma}_1^c = 44.7\%$); and least likely to give when P2 can give but decides to keep in the nonexclusive game ($\hat{\gamma}_{1K}^n = 23.3\%$). 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 holds income effects, distributional preferences, and social image concerns constant across games, enabling us to rule out these alternative explanations.

We next compare the behavior of Initial Movers (P2 in the treatment games and P1 in the control game). This allows us to investigate whether Initial Movers are motivated by the possibility that their beneficiary will pay forward their generosity to magnify their impact. If so, we would expect giving to be greater among P2 in the treatment game than P1 in the control game. However, we find the opposite pattern. P1's giving rate in the control game is 44.7%, which is significantly greater than P2's giving rates of 40.2% and 39.0% in the exclusive and nonexclusive games, respectively ($p < 0.05$). It appears that expectations about P1 paying forward P2's generosity play a negligible role in guiding P2's giving decision.

How can we explain these behaviors? We embed altruism, inequity aversion, and indirect reciprocity incentives as psychological components in a game-theoretic framework that extends Dufwenberg and Kirchsteiger (2004), while incorporating elements from Fehr and Schmidt (1999) and Rabin (1993).[2] We turn on or shut off each of the three psychological components, generating predictions on binary giving decisions under $2^3 = 8$ utility specifications. Our models include (i) the standard model when all factors are turned off; (ii) inequity aversion as formalized by Fehr and Schmidt (1999); and (iii) the modification of Dufwenberg and Kirchsteiger (2004)

---

[1] These multiplier methods are commonly used to represent the positive externalities that result when individuals engage in kind acts. For example, the act of donating a kidney to a stranger demonstrates that paired-kidney exchange can work, promoting public confidence in the allocation mechanism and motivating greater organ donation rates for future patients. These benefits accrue to society overall, beyond the private benefit to the organ recipient.

[2] We focus on upstream indirect reciprocity: Receiving past kindness motivates an agent to help a third party (Mujcic and Leibbrandt, 2018; van Apeldoorn and Schram, 2016). This form of indirect reciprocity is less studied compared to downstream indirect reciprocity, in which an agent is more likely to *receive* help after helping another (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). Downstream indirect reciprocity focuses less on the psychological motivation of the potential helper and more on the reputation of the potential recipient of help, making it outside the scope of our paper.

with indirect reciprocity motives rather than direct reciprocity motives. We then assess the explanatory power of each model by comparing its predictions with subjects' behaviors in the experiment. These variations enable us to quantify the empirical importance of altruism, reciprocity, and inequity aversion in explaining experimental behavior.

We find that the most general model with altruism, reciprocity, and inequity aversion explains the behavior of 90% of subjects. Altruism and reciprocity are key to explaining why people are more likely to give after having received a gift: The model excluding altruism can only explain the behavior of 30% of subjects, while the model excluding reciprocity can only explain the behavior of 70% of subjects. In contrast, inequity aversion plays a marginal role and only helps explain why P2's giving does *not* rise in situations where P1 could pay her generosity forward. The model that excludes inequity aversion performs almost as well as the full model, in that it explains the behavior of 88% of subjects.

Overall, altruism and indirect reciprocity have high explanatory power over pay-it-forward behavior, and can explain why chains of generosity propagate once started. However, our subjects were not motivated to give more based on the possibility that their generosity would be paid forward. Our experimental evidence suggests that chains of generosity are difficult to start.

Our contributions are threefold. First, we introduce a simple, novel experiment that establishes the role of indirect reciprocity motives in pay-it-forward behavior while controlling for alternative explanations. Many prior papers fail to determine if reciprocity intentions truly motivate pay-it-forward behavior, since they cannot rule out the income effect, where the act of receiving a gift itself can make subjects more likely to give through increasing their wealth (Herne et al., 2013; van Apeldoorn and Schram, 2016; Simpson et al., 2018; Mujcic and Leibbrandt, 2018; Attanasi et al., 2019). Furthermore, to 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. In addition, subjects participate in our experiment online without observation or communication from other subjects. This affords some control over social image considerations, which have been established to be important in reciprocal interactions, both theoretically (Charness and Rabin, 2002; Sobel, 2005; Cox et al., 2008; Battigalli and Dufwenberg, 2022) and empirically (Malmendier et al., 2014).

Second, most game-theoretic frameworks focus on direct reciprocity between two individuals who directly interact (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Seinen and Schram, 2006; Cox et al., 2007; Battigalli and Dufwenberg, 2009; Berger, 2011; Gong and Yang, 2019; Gaudeul et al., 2021; Battigalli and Dufwenberg, 2022), but cooperative communities frequently involve three or more individuals who do not directly interact.[3] To our knowledge, our paper is the first to develop a behavioral game-theoretic framework for the systematic investigation of indirect reciprocity, promoting reciprocal exchange in environments with indirect interaction. We find that altruism and reciprocity incentives can explain why kindness begets further kindness. By promoting the propagation of generosity, they help sustain chains of giving once started. Inequity aversion, however, presents a barrier to starting these chains. Our paper has important ramifications for how to foster cultures of cooperation within workplaces, neighborhoods, and communities.

Third, our theory and experiment complement each other in exploring the roles of type-based, outcome-based, and intentions-based models of fairness on pay-it-forward behavior. 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 prosocial parameters (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). We combine elements from these models to evaluate the importance of each in rationalizing subjects' behavior. Specifically, type-based components of our model allow us to fix subjects' structural prosocial parameters across various nodes of the game. Eliciting subjects' full strategy sets then enables us to perform within-subject comparisons across different nodes of the game. Moreover, our within-subject design allows us to explicitly quantify the proportion of subjects whose strategies align with theoretical predictions, following methods in Charness and Rabin (2002).

The rest of the paper is organized as follows. Section 2 introduces the psychological game-theoretic framework and derives predicted equilibrium giving rates. Section 3 describes the experimental procedure. Section 4 compares our experimental results with theoretical predictions and evaluates the roles of altruism, inequity aversion, and reciprocity in rationalizing experimental behavior. Section 5 concludes. The appendix collects omitted proofs and experimental details.

2. Theory

2.1. Model

We construct a model tailored to our experiment to better understand subjects' giving decisions in the games summarized in Fig. 1. All the games we consider are finite-action multistage games with observed actions, perfect recall, and without moves of nature. 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. We follow the framework of Dufwenberg and Kirchsteiger (2004) but make three modifications. First, we incorporate an altruistic payoff component. Second, we incorporate indirect reciprocity rather than direct reciprocity. Third, we incorporate inequity aversion from Fehr and Schmidt (1999), and make adjustments to account for expected payoffs.

---

[3] Wu (2018) and Jiang and Wu (2019) discuss models involving indirect interactions among more than two players. Reciprocal behavior has also been investigated in 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).

Let $N = \{0, 1, \ldots, n\}$ denote the set of $n + 1$ 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 *non-terminal* histories of a game. In multistage games with observed actions, H is the domain of pure strategies as well as the domain of behavior strategies. Let S_i denote player i's pure strategy set, and $S = \times_{i \in N} S_i$ the set of pure strategy profiles. Let $C_i(h)$ denote the deterministic choices available at history h for player i, and $\Delta(C_i(h))$ the set of randomized choices available at history h for player i. The set of behavior strategies of player $i \in N$ is denoted by $\Sigma_i = \prod_{h \in H} \Delta(C_i(h))$, 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 behavior strategy profiles σ of all players, and $\Sigma_{-i} = \prod_{j \in N \setminus \{i\}} \Sigma_j$ the collection of behavior strategy profiles σ_{-i} of all players other than i. Let Σ'_{ij} be the set of beliefs of player i about the behavior strategy of player j (i.e., i's first-order beliefs).[4]

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). We only consider Dirac beliefs about the beliefs of others, so $\Sigma'_{ij} = \Sigma_j$ and $\Sigma''_{ijk} = \Sigma'_{jk} = \Sigma_k$.

Define H_i to be the set in which player i is "active," namely, the set of choices available to them has a cardinality of 2 or greater. If $|C_i(h')| = 1$ for some history h', we interpret it as player i having just one available action, namely, to "wait." With $\sigma_i \in \Sigma_i$ and $h \in H_i$, let $\sigma_i^{[h]}$ denote the behavior strategy that prescribes the same mixed actions as σ_i except for the choices made by i at history h, which are assigned probability 1. 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 $\sigma'^{[h]}_{ij}$ and $\sigma''^{[h]}_{ijk}$ analogously.

Player i's utility depends on behavior strategy profile σ, first-order belief profile σ', and second-order belief profile σ'', which we summarize by a vector $\vec{\sigma} \equiv (\sigma, \sigma', \sigma'')$. These profiles of strategies and beliefs, in turn, determine the expected payoffs of players, which comprise her expected material payoff and three psychological components: (i) expected altruistic payoff, (ii) expected indirect reciprocity payoff, and (iii) expected inequity aversion payoff.

The **altruistic component** is a function of i's altruistic factor A_i and the expected material payoffs of other players regardless of their distribution:

$$A_i \sum_{j \neq i} \pi_j^h \left(\sigma_i^{[h]}, \left(\sigma'^{[h]}_{ik} \right)_{k \neq i} \right).$$

The **indirect reciprocity component** depends on i's indirect reciprocity parameter Z_i, player i's kindness to player j, $\kappa_{ij}^h : \Sigma_i \times \Sigma'_{ij} \to \mathbb{R}$, and player i's belief of player k's kindness to player i, $\lambda_{iki}^h : \Sigma'_{ik} \times \Sigma''_{ikl} \to \mathbb{R}$:

$$Z_i \lambda_{iki}^h \left(\sigma'^{[h]}_{ik}, \left(\sigma''^{[h]}_{ikl} \right)_{l \neq k} \right) \kappa_{ij}^h \left(\sigma_i^{[h]}, \left(\sigma'^{[h]}_{ik} \right)_{k \neq i} \right).$$

Both κ_{ij}^h and λ_{ikl}^h are adapted from Dufwenberg and Kirchsteiger (2004), where we use an indirect reciprocity component rather than a direct reciprocity component since our games only allow P1 to reciprocate P2's kindness by helping P0, rather than directly reciprocating P2.

To be more precise, player i's kindness to player j is positive (negative) if he chooses an action that gives player j a strictly higher (lower) expected payoff than his equitable payoff, $\pi_j^{h,Q_i} : \Sigma_{ik} \to \mathbb{R}$, which is given by the average of player j's lowest and highest possible material payoff based on i's strategy. In our games, only P1 can indirectly reciprocate. In particular, P1's kindness to P0 at history h is captured by the function $\kappa_{10}^h : \Sigma_1 \times \Sigma'_2 \to \mathbb{R}$ defined as:

$$\kappa_{10}^h \left(\sigma_1^{[h]}, \sigma'^{[h]}_2 \right) = \pi_0^h \left(\sigma_1^{[h]}, \sigma'^{[h]}_2 \right) - \pi_0^{h,Q_1} \left(\sigma'^{[h]}_2 \right),$$

where

$$\pi_0^{h,Q_1} \left(\left(\sigma'^{[h]}_2 \right)_{k \neq i} \right) = \frac{1}{2} \left[\max_{c_1 \in C_1(h)} \pi_0^h \left(\widetilde{\sigma}_1^h \left(\sigma_1^{[h]}, c_1 \right), \sigma'^{[h]}_2 \right) + \min_{c_1 \in C_1(h)} \pi_0^h \left(\widetilde{\sigma}_1^h \left(\sigma_1^{[h]}, c_1 \right), \sigma'^{[h]}_2 \right) \right]$$

is P0's *equitable payoff* with respect to P1. $\widetilde{\sigma}_1^h(\sigma_1^{[h]}, c_1)$ is the modified strategy such that $\widetilde{\sigma}_1^h(\sigma_1^{[h]}, c_1)(h') = \sigma_i^{[h]}(h')$ for each $h' \in H \setminus \{h\}$, and $\widetilde{\sigma}_1^h(\sigma_1^{[h]}, c_1)(h) = c_i$.

The last component of indirect reciprocity, player i's belief of player k's kindness to him, is given by the difference between i's expected payoff given k's action and i's equitable payoff given i's belief of k's belief about other players' strategies. In our games, we are only interested in P1's belief of P2's kindness to him at history h, given by:

$$\lambda_{121}^h \left(\sigma'^{[h]}_{12}, \sigma'^{[h]}_{121} \right) = \pi_1^h \left(\sigma'^{[h]}_{12}, \sigma''^{[h]}_{121} \right) - \pi_1^{h,Q_2} \left(\sigma''^{[h]}_{121} \right).$$

The **inequity aversion component** depends on inequity aversion parameters α_i and β_i, incorporated from Fehr and Schmidt (1999). Player i receives disutility α_i for each unit of payoff lower than others (*disadvantageous inequity aversion*) and disutility β_i for

---

[4] We allow for randomized choices since they help ensure the existence of equilibrium in finite games (Chapter 6, Theorem 12, Battigalli et al., 2025). Randomized choices enable modeling of optimal strategies in experiments such as ours, where subjects are randomly matched in groups and form beliefs about others' actions. Note, however, that expected utility maximizers do not have a strict incentive to randomize, since for any conjecture about the behaviors of others, for any mixed action, there exists a pure action that would yield the same or higher expected payoff (Battigalli et al., 2025).

each unit of payoff higher than others (*advantageous inequity aversion*), where $\alpha_i \geq \beta_i$ and $0 \leq \beta_i \leq 1$. The disadvantageous inequity aversion payoff is given by

$$\alpha_i \frac{1}{n} \sum_{j \neq i} \max \left\{ \pi_j^h \left(s_i^{[h]}, \left(\sigma_{ik}^{[h]} \right)_{k \neq i} \right) - \pi_i^h \left(s_i^{[h]}, \left(\sigma_{ik}^{[h]} \right)_{k \neq i} \right), 0 \right\},$$

while the advantageous inequity aversion by

$$\beta_i \frac{1}{n} \sum_{j \neq i} \max \left\{ \pi_i^h \left(s_i^{[h]}, \left(\sigma_{ik}^{[h]} \right)_{k \neq i} \right) - \pi_j^h \left(s_i^{[h]}, \left(\sigma_{ik}^{[h]} \right)_{k \neq i} \right), 0 \right\}.$$

Player i's expected utility at history h takes the following general form (which we will simplify for our games):

$$u_i^h \left(\sigma_i, \left(\sigma_{ij}' \right)_{j \neq i}, \left(\sigma_{ijk}'' \right)_{j \neq i, j \neq k} \right) \tag{1}$$

$$= \underbrace{\pi_i^h \left(\sigma_i^{[h]}, \left(\sigma_{ij}'^{[h]} \right)_{j \neq i} \right)}_{\text{expected material payoff}} + \underbrace{A_i \sum_{j \neq i} \pi_j^h \left(\sigma_i^{[h]}, \left(\sigma_{ik}'^{[h]} \right)_{k \neq i} \right)}_{\text{expected altruistic payoff}}$$

$$+ \underbrace{\sum_{j \neq i} \sum_{k \notin \{i,j\}} Z_i \lambda_{iki}^h \left(\sigma_{ik}'^{[h]}, \left(\sigma_{ikl}''^{[h]} \right)_{l \neq k} \right) \kappa_{ij}^h \left(\sigma_i^{[h]}, \left(\sigma_{ik}'^{[h]} \right)_{k \neq i} \right)}_{\text{expected indirect reciprocity payoff}}$$

$$- \sum_{s \in S} P_\sigma^h(s) \left[\overbrace{\underbrace{\alpha_i \frac{1}{n} \sum_{j \neq i} \max \left\{ \pi_j^h \left(s_i^{[h]}, \left(\sigma_{ik}'^{[h]} \right)_{k \neq i} \right) - \pi_i^h \left(s_i^{[h]}, \left(\sigma_{ik}'^{[h]} \right)_{k \neq i} \right), 0 \right\}}_{\text{expected advantageous inequity aversion payoff}}}^{\text{expected disadvantageous inequity aversion payoff}} \right.$$

$$\left. + \underbrace{\beta_i \frac{1}{n} \sum_{j \neq i} \max \left\{ \pi_i^h \left(s_i^{[h]}, \left(\sigma_{ik}'^{[h]} \right)_{k \neq i} \right) - \pi_j^h \left(s_i^{[h]}, \left(\sigma_{ik}'^{[h]} \right)_{k \neq i} \right), 0 \right\}}_{\text{}} \right].$$

$$\underbrace{}_{\text{expected inequity aversion payoff}}$$

To clarify, $\pi_i^h \left(\sigma_i^{[h]}, \left(\sigma_{ij}^{[h]} \right)_{j \neq i} \right)$ is i's expected material payoff at history h given $\left(\sigma_i^{[h]}, \left(\sigma_{ij}^{[h]} \right)_{j \neq i} \right)$ and $P_\sigma^h(s)$ is the probability of s according to σ computed by means of the transformation from behavior to mixed strategies by Kuhn (1953) that preserves the probabilities of terminal histories and hence preserves expected payoffs. Namely, for all $s \in \prod_{i \in N} \prod_{h \in H} C_i(h)$, $P_\sigma^h(s) = \prod_{i \in N} \prod_{h \in H} \sigma_{i,h}(s_{i,h})$.

In our games, only P1 can indirectly reciprocate, so only Z_1 needs to be considered. P1's expected utility at history $h \in H$ takes the following form:

$$u_1^h \left(\sigma_1, \sigma_{12}', \sigma_{121}'' \right) = \underbrace{\pi_1^h \left(\sigma_1^{[h]}, \sigma_{12}'^{[h]} \right)}_{\text{expected material payoff}} + \underbrace{A_1 \sum_{j \in \{0,2\}} \pi_j^h \left(\sigma_1^{[h]}, \sigma_{12}'^{[h]} \right)}_{\text{expected altruistic payoff}} \tag{2}$$

$$+ \underbrace{Z_1 \lambda_{121}^h \left(\sigma_{12}'^{[h]}, \sigma_{121}''^{[h]} \right) \kappa_{10}^h \left(\sigma_1^{[h]}, \sigma_{12}'^{[h]} \right)}_{\text{expected indirect reciprocity payoff}}$$

$$- \sum_{s \in S} P_\sigma^h(s) \left[\overbrace{\underbrace{\alpha_1 \frac{1}{2} \sum_{j \in \{0,2\}} \max \left\{ \pi_j^h \left(s_1^{[h]}, \sigma_{12}'^{[h]} \right) - \pi_1^h \left(s_1^{[h]}, \sigma_{12}'^{[h]} \right), 0 \right\}}_{\text{expected advantageous inequity aversion payoff}}}^{\text{expected disadvantageous inequity aversion payoff}} \right.$$

$$\left. + \underbrace{\beta_1 \frac{1}{2} \sum_{j \in \{0,2\}} \max \left\{ \pi_1^h \left(s_1^{[h]}, \sigma_{12}'^{[h]} \right) - \pi_j^h \left(s_1^{[h]}, \sigma_{12}'^{[h]} \right), 0 \right\}}_{\text{}} \right],$$

$$\underbrace{}_{\text{expected inequity aversion payoff}}$$

and P2's expected utility at the beginning of the game (the only instance when she is called upon to act) takes the following form:

$$u_2^h\left(\sigma_2,\sigma_{21}'\right)=\underbrace{\pi_2^h\left(\sigma_2^{[h]},\sigma_{21}'^{[h]}\right)}_{\text{expected material payoff}}+\underbrace{A_2\sum_{j\in\{0,1\}}\pi_j^h\left(\sigma_2^{[h]},\sigma_{21}'^{[h]}\right)}_{\text{expected altruistic payoff}}\qquad(3)$$

$$-\sum_{s\in S}P_\sigma^h(s)\Bigg[\overbrace{\alpha_2\frac{1}{2}\sum_{j\in\{0,1\}}\max\left\{\pi_j^h\left(s_2^{[h]},\sigma_{21}'^{[h]}\right)-\pi_2^h\left(s_2^{[h]},\sigma_{21}'^{[h]}\right),0\right\}}^{\text{expected disadvantageous inequity aversion payoff}}$$

$$\underbrace{+\beta_2\frac{1}{2}\sum_{j\in\{0,1\}}\max\left\{\pi_2^h\left(s_2^{[h]},\sigma_{21}'^{[h]}\right)-\pi_j^h\left(s_2^{[h]},\sigma_{21}'^{[h]}\right),0\right\}}_{\text{expected advantageous inequity aversion payoff}}\Bigg].$$

expected inequity aversion payoff

Both κ_{10}^h and λ_{121}^h are modified from Dufwenberg and Kirchsteiger (2004), where we use an indirect reciprocity component rather than a direct reciprocity component since our games only allow P1 to reciprocate P2's kindness by helping P0, rather than directly reciprocating P2. To be more precise, κ_{10}^h depends only on $\left(\sigma_1^{[h]},\sigma_{12}'^{[h]}\right)$ and λ_{121}^h depends only on $(\sigma_{12}'^{[h]},\sigma_{121}''^{[h]})$.

Next, we define $2^3=8$ specifications of the utility function in which none, some, or all of the three psychological components—altruism, indirect reciprocity, and inequity aversion—are assumed away. We formulate key predictions based on the utility specifications and test them against our experimental results in Section 4.

Definition 1. *Standard/selfish (S) utility* ignores psychological components and assumes that all parameters are zero: $A_i=0$, $\alpha_i=0$, $\beta_i=0$, and $Z_i=0$ for all i.

Altruistic (A), *Inequity-averse (I)*, and *Reciprocal (R)* utilities respectively assume that the relevant parameters are positive (namely, $A_i>0$ for all i; $\alpha_i>0$ and $\beta_i>0$ for all i; and $Z_i>0$ for all i, respectively) and other parameters are zero.

Altruistic and Inequity-averse (AI), *Altruistic and Reciprocal (AR)*, as well as *Inequity-averse and Reciprocal (IR)* utilities respectively assume that the relevant sets of parameters are positive (namely, $A_i>0$, $\alpha_i>0$, and $\beta_i>0$ for all i; $A_i>0$ and $Z_i>0$ for all i; and $\alpha_i>0$, $\beta_i>0$, and $Z_i>0$, respectively) and other parameters are zero.

Altruistic, Inequity-averse, and Reciprocal (AIR) utility assumes that all four parameters are positive: $A_i>0$, $\alpha_i>0$, $\beta_i>0$, and $Z_i>0$ for all i.

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 constitute a *dynamic reciprocity equilibrium* if and only if (i) (consistency) players have correct beliefs about other players' behavior strategies, i.e., $\sigma=\sigma_i'=\sigma_{ij}''$ for any players i and j[5]; and (ii) (utility maximization) for each player i, strategy profile σ_i maximizes player i's utility at each information set given first-order and second-order beliefs σ_i' and σ_{ij}''.

Theorem 1. *A dynamic reciprocity equilibrium always exists.*[6]
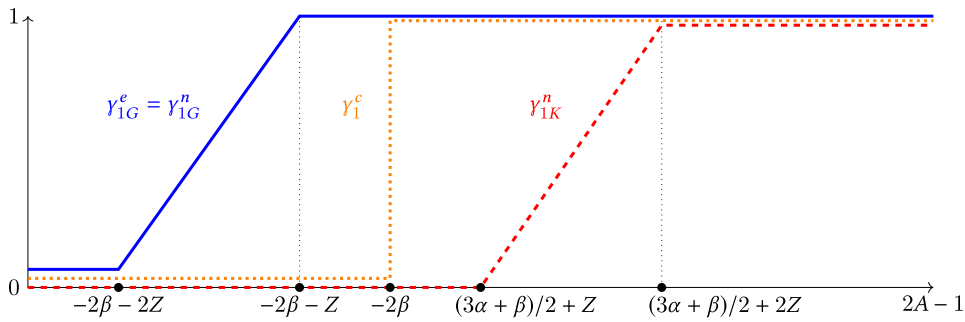
Proof. See Appendix A.2. □

We elicit full strategy sets and beliefs in our experiment (see the experimental procedure in Section 3 regarding the elicitation of strategy sets and Appendix B.1 on the elicitation of beliefs). The model specifies that the key belief parameters are first-order and second-order beliefs about P1's likelihood of giving to P0. Appendix B.1 shows that elicited beliefs match empirical giving rates, supporting the consistency condition in our definition of equilibrium. In the rest of the paper, we evaluate theoretical and empirical behaviors, assuming subjects play under dynamic reciprocity equilibrium.

2.2. Giving decisions

There are three players—P2, P1, and P0—in each of our three games. Each player's index denotes the number of players behind her in the chain: P0 is the last potential recipient, P1 is the last player to decide on giving, and P2, if allowed, is the penultimate player to decide on giving. The game trees are depicted in Fig. 1. To simplify the exposition of giving rates, we introduce the following notation.

---

[5] The consistency assumption is a stricter version of the coherence assumption specified by Battigalli et al. (2019) and Battigalli and Dufwenberg (2022), which requires that each player's first-order beliefs be a marginalization of their second-order beliefs.

[6] In general, a dynamic reciprocity equilibrium is not necessarily unique. However in our games, 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.

Note: The figure depicts the equilibrium probabilities of giving by Last Movers as $2A - 1$ increases.

Fig. 2. Last Movers' equilibrium giving rates comparisons.

Definition 3. For any real number x, 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 (Fig. 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 gives if and only if $2A_1 + 2\beta_1 \geq 1$.*

P1 can either keep all 3 chips (so that P0 gains no chips) or give 1 of them to P0 (so that P0 gains 2 chips). When P1 gives 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 $2A_1$ 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 $2\beta_1$ from giving and equalizing payoffs across all three players. Overall, the psychological gain of giving by P1 is

$$2A_1 + 2\beta_1. \tag{4}$$

Fig. 2 depicts P1's equilibrium giving rate as A_1 varies. In equilibrium, the giving rate is either 0 or 1 except for when $2A_1 + 2\beta_1 = 1$, so it can be represented by an indicator function: $\gamma_1^c = \mathbf{1}_{2A_1 + 2\beta_1 \geq 1}$.[7] P1 is more inclined to give the higher her altruistic factor A_1 and advantageous inequity aversion β_1 (that is, the more she dislikes having more than other players). Pure altruism A_1 and/or advantageous inequity aversion β_1—but not disadvantageous inequity aversion α_1 or reciprocity Z_1—helps rationalize giving by P1 in the control game. Note that reciprocity does not play a role here because only one player takes an action, and others do not have the opportunity to reciprocate.
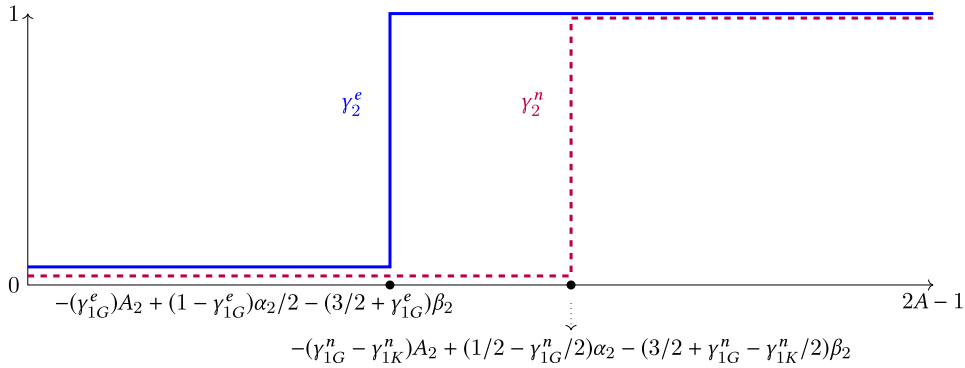
2.2.2. Exclusive game

Second, consider a three-player game in which P1 can only give to P0 if P2 gave to P1 first (Fig. 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 any equilibrium of the exclusive game, P2 gives if $1 + \alpha_2 \frac{1 - \gamma_{1G}^e}{2} \leq A_2(2 + \gamma_{1G}^e) + \beta_2 \frac{3 + 2\gamma_{1G}^e}{2}$, and P1 gives with probability $\gamma_{1G}^e = [\![\frac{2A_1 + 2\beta_1 - 1}{Z_1} + 2]\!]$.*

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 gave, P2 incurs utility loss from disadvantageous inequity of $(3 - 2)\alpha_2/2 = \alpha_2/2$. Because P1 keeps with probability $1 - \gamma_{1G}^e$ after P2 gave, giving would lower P2's expected utility by $(1 - \gamma_{1G}^e)\alpha_2/2$.

The right-hand side of the inequality represents the two ways P2 gains from giving. Giving increases P2's altruistic payoff by $A_2(2 + \gamma_{1G}^e)$ in expectation. Giving also rectifies inequity aversion, in that P2 is less likely to have more than other players. If P2 keeps,

---

[7] When $2A_1 + 2\beta_1 = 1$, P1 is indifferent between giving and keeping, since the change in material payoffs equals 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.

Note: The figure depicts the equilibrium probabilities of giving by Initial Movers as $2A - 1$ increases.

Fig. 3. Initial Movers' equilibrium giving rates.

she suffers disutility of $[(3 - 1) + (3 - 0)]\beta_2/2 = 5\beta_2/2$ 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 $1 - \gamma_{1G}^e$), then P2 suffers $(2 - 0)\beta_2/2 = \beta_2$ from getting more than P0. Hence, by giving, P2 gains in expectation $5\beta_2/2 - (1 - \gamma_{1G}^e)\beta_2 = (3/2 + \gamma_{1G}^e)\beta_2$.

In summary, altruism motivates P2 to give. However, the effect of inequity aversion is ambiguous. Advantageous inequity aversion motivates P2 to give since she loses utility from having more than the other two players. Disadvantageous inequity aversion motivates P2 to keep since giving to P1 could lead her to end up with less than P1.

Next, we consider P1's strategy. 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 $2A_1$ and $2\beta_1$ from equalizing payoffs. Furthermore, P1 earns an indirect reciprocity payoff of $Z_1(2 - \gamma_{1G}'')$, 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 gave is

$$2A_1 + 2\beta_1 + Z_1(2 - \gamma_{1G}''). \tag{5}$$

In equilibrium, P1's second-order belief must equate with her strategy ($\gamma_{1G}'' = \gamma_{1G}^e$). If $2A_1 + 2\beta_1 + Z_1 \geq 1$, then $\gamma_{1G}^e = 1$; if $2A_1 + 2\beta_1 + 2Z_1 \leq 1$, then $\gamma_{1G}^e = 0$. Otherwise, a mixed strategy is needed to equate the equilibrium strategy of P1 and the belief of P2: P1 gives with a probability strictly between 0 and 1 that makes her indifferent between giving and keeping. P1's inclination to give increases with altruism A_1, advantageous inequity aversion β_1, and indirect reciprocity Z_1. Fig. 2 depicts how P1's equilibrium giving rate varies with altruism A_1. Fig. 3 depicts how P2's equilibrium giving rate varies with A_2.

2.2.3. Nonexclusive game

Finally, consider a three-player game in which P0's channel of receiving chips is nonexclusive (Fig. 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 any equilibrium of the nonexclusive game, P2 gives if and only if* $1 + \alpha_2 \frac{1 - \gamma_{1G}^n}{2} \leq A_2(2 + \gamma_{1G}^n - \gamma_{1K}^n) + \beta_2 \frac{3 + 2\gamma_{1G}^n - \gamma_{1K}^n}{2}$, *P1 gives with probability* $\gamma_{1G}^n = [\![\frac{2A_1 + 2\beta_1 - 1}{Z_1} + 2]\!]$ *after P2 gave, and P1 gives with probability* $\gamma_{1K}^n = [\![\frac{4A_1 - 3\alpha_1 - \beta_1 - 2}{2Z_1} - 1]\!]$ *after P2 kept.*

P2's gains from giving are similar to those in the exclusive game. The only difference is that after P2 kept, P1 gives 1 chip with probability γ_{1K}^n, 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 in the exclusive game by $\gamma_{1K}^n(A_2 + \beta_2/2)$.

Finally, we consider P1's strategy. If P2 chooses to give, then P1 faces the same trade-off as in the exclusive game. Therefore, the equilibrium giving rate γ_{1G}^n 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 $Z_1(2 - \gamma_{1K}'')$. 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 an increase of $3\alpha_1/2$ units in disadvantageous inequity aversion and an increase of $\beta_1/2$ units in advantageous inequity aversion. Overall, the psychological gain of giving by P1 after P2 keeps is

$$2A_1 - 3\alpha_1/2 - \beta_1/2 - Z_1(2 - \gamma_{1G}'') = 2A_1 + 2\beta_1 - 3\alpha_1/2 - 3\beta_1/2 - Z_1(2 - \gamma_{1G}''). \tag{6}$$

Table 1

Summary of conditions for positive equilibrium giving rate ($\gamma > 0$).

Last Movers (P1 in all games)	
P1 in the control game	$2A_1 + 2\beta_1 > 1$
P1 in the exclusive game after P2 gave	$2A_1 + 2\beta_1 + 2Z_1 > 1$
P1 in the nonexclusive game after P2 gave	$2A_1 + 2\beta_1 + 2Z_1 > 1$
P1 in the nonexclusive game after P2 kept	$2A_1 + 2\beta_1 + 2Z_1 - 3\alpha_1 - 3\beta_1 - 3Z_1 > 1$

Initial Movers (P1 in the control game and P2 in the treatment games)	
P1 in the control game	$2A_1 + 2\beta_1 > 1$
P2 in the exclusive game	$2A_2 + 2\beta_2 + \gamma_{1G}^e A_2 - \alpha_2 \frac{1-\gamma_{1G}^e}{2} - \beta_2 \frac{1-2\gamma_{1G}^e}{2} > 1$
P2 in the nonexclusive game	$2A_2 + 2\beta_2 + (\gamma_{1G}^n - \gamma_{1K}^n)A_2 - \alpha_2 \frac{1-\gamma_{1G}^n}{2} - \beta_2 \frac{1-2\gamma_{1G}^n+\gamma_{1K}^n}{2} > 1$

Note: The table summarizes the conditions for positive equilibrium rates, which are deduced from the three lemmas by setting $\gamma > 0$.

Table 2

Comparisons of giving: theoretical predictions.

Last Movers (P1 in all games)								
Model \ Comparison	S	A	I	R	AI	IR	AR	AIR
1: γ_{1G}^n versus γ_{1K}^n	$\gamma_{1G}^n \sim \gamma_{1K}^n \sim 0$	$\gamma_{1G}^n \sim \gamma_{1K}^n > 0$	$\gamma_{1G}^n > \gamma_{1K}^n \sim 0$	$\gamma_{1G}^n \sim \gamma_{1K}^n \sim 0$	$\gamma_{1G}^n > \gamma_{1K}^n > 0$	$\gamma_{1G}^n > \gamma_{1K}^n \sim 0$	$\gamma_{1G}^n > \gamma_{1K}^n > 0$	$\gamma_{1G}^n > \gamma_{1K}^n > 0$
2: γ_{1G}^e versus γ_1^c	$\gamma_{1G}^e \sim \gamma_1^c \sim 0$	$\gamma_{1G}^e \sim \gamma_1^c > 0$	$\gamma_{1G}^e \sim \gamma_1^c > 0$	$\gamma_{1G}^e > \gamma_1^c \sim 0$	$\gamma_{1G}^e \sim \gamma_1^c > 0$	$\gamma_{1G}^e > \gamma_1^c > 0$	$\gamma_{1G}^e > \gamma_1^c > 0$	$\gamma_{1G}^e > \gamma_1^c > 0$
3: γ_{1G}^n versus γ_1^c	$\gamma_{1G}^n \sim \gamma_1^c \sim 0$	$\gamma_{1G}^n \sim \gamma_1^c > 0$	$\gamma_{1G}^n > \gamma_1^c > 0$	$\gamma_{1G}^n > \gamma_1^c \sim 0$	$\gamma_{1G}^n > \gamma_1^c > 0$	$\gamma_{1G}^n > \gamma_1^c > 0$	$\gamma_{1G}^n > \gamma_1^c > 0$	$\gamma_{1G}^n > \gamma_1^c > 0$
4: γ_1^c versus γ_{1K}^n	$\gamma_1^c \sim \gamma_{1K}^n \sim 0$	$\gamma_1^c \sim \gamma_{1K}^n > 0$	$\gamma_1^c > \gamma_{1K}^n \sim 0$	$\gamma_1^c \sim \gamma_{1K}^n \sim 0$	$\gamma_1^c > \gamma_{1K}^n \sim 0$	$\gamma_1^c > \gamma_{1K}^n \sim 0$	$\gamma_1^c > \gamma_{1K}^n > 0$	$\gamma_1^c > \gamma_{1K}^n > 0$
5: γ_{1G}^n versus γ_{1G}^e	$\gamma_{1G}^n \sim \gamma_{1G}^e \sim 0$	$\gamma_{1G}^n \sim \gamma_{1G}^e > 0$	$\gamma_{1G}^n \sim \gamma_{1G}^e > 0$	$\gamma_{1G}^n \sim \gamma_{1G}^e > 0$	$\gamma_{1G}^n \sim \gamma_{1G}^e > 0$	$\gamma_{1G}^n \sim \gamma_{1G}^e > 0$	$\gamma_{1G}^n \sim \gamma_{1G}^e > 0$	$\gamma_{1G}^n \sim \gamma_{1G}^e > 0$

Initial Movers (P1 in the control game and P2 in the treatment games)								
Model \ Comparison	S	A	I	R	AI	IR	AR	AIR
6: γ_2^e versus γ_2^n	$\gamma_2^e \sim \gamma_2^n \sim 0$	$\gamma_2^e > \gamma_2^n > 0$	$\gamma_2^e \sim \gamma_2^n > 0$	$\gamma_2^e \sim \gamma_2^n \sim 0$	$\gamma_2^e > \gamma_2^n > 0$	$\gamma_2^e \sim \gamma_2^n > 0$	$\gamma_2^e > \gamma_2^n > 0$	$\gamma_2^e > \gamma_2^n > 0$
7: γ_2^e versus γ_1^c	$\gamma_2^e \sim \gamma_1^c \sim 0$	$\gamma_2^e > \gamma_1^c > 0$	depends	$\gamma_2^e \sim \gamma_1^c \sim 0$	depends	depends	$\gamma_2^e > \gamma_1^c > 0$	depends
8: γ_2^n versus γ_1^c	$\gamma_2^n \sim \gamma_1^c \sim 0$	$\gamma_2^n > \gamma_1^c > 0$	depends	$\gamma_2^n \sim \gamma_1^c \sim 0$	depends	depends	$\gamma_2^n > \gamma_1^c > 0$	depends

Note: 'depends' indicates the predicted comparison depends on inequity aversion parameters α_i and β_i.

Labels: Giving rate is denoted by γ. The superscript 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. S - standard model; A - altruism; R - reciprocity; I - inequity aversion.

Note that the equilibrium probabilities of giving after P1 gave and after P1 kept cannot be both strictly between 0 and 1 under any combination of parameters. This is because the conditions for indifference differ at the two decision nodes. When P1 is indifferent between giving and keeping at one decision node, she is not indifferent at the other.

Since we are interested in characterizing when a player may give, we summarize in Table 1 the conditions under which giving is part of the equilibrium strategy (i.e., $\gamma > 0$).

2.3. Comparisons of giving: theoretical predictions

Our theory and experiment complement each other in generating and testing predictions of giving strategies. In the experiment, we elicit subjects' giving decisions at all nodes of all games using the strategy method. This enables us to compare behavior at different nodes within subjects. To make these comparisons, we say that one is more inclined to give in the following sense.

Definition 4. Player i is **more inclined** to take action s at node h than player j to take action s' at node h', $\sigma_i(s|h) \succ \sigma_j(s'|h')$, if given $(A_i, Z_i, \alpha_i, \beta_i) = (A_j, Z_j, \alpha_j, \beta_j)$, in equilibrium, $\sigma_i(s|h) \geq \sigma_j(s'|h')$ for all parameters, and the inequality holds strictly for some parameters. Player i is **equally inclined** to take action s at node h as player j to take action s' at node h', $\sigma_i(s|h) \sim \sigma_j(s'|h')$, if given $(A_i, Z_i, \alpha_i, \beta_i) = (A_j, Z_j, \alpha_j, \beta_j)$, in equilibrium, $\sigma_i(s|h) = \sigma_j(s'|h')$ for all parameters.

Note that the comparison could be between two choices of the same player in the same information set or in different information sets, and it could also be between two choices of different players.

Table 2 summarizes pairwise comparisons of giving. These comparisons are formally presented as propositions in Appendix A under each of the eight utility functions ("models") listed in Definition 1, which turn on or off each of the three psychological components we consider. These propositions generate one prediction for each comparison under each model. In the exposition below, the order of comparisons derives from the order in which the propositions must be proven (see Appendix A).

First, note that a standard model without the aforementioned psychological components (Model S) predicts no giving by any player under any circumstance, since giving strictly decreases subjects' material payoffs. For all other models, we first consider predictions

for Last Movers: P1 in all games. We then consider predictions for Initial Movers: P2 in the treatment games and P1 in the control game.

2.3.1. Predictions for Last Movers

Rows 1–5 of Table 2 report predicted pairwise comparisons of giving rates of Last Movers (P1 in all games).[8] In the most general utility specification with altruism, reciprocity, and inequity aversion (Model AIR), the giving rates by the Last Mover are ordered $\gamma_{1G}^e \sim \gamma_{1G}^n > \gamma_1^c > \gamma_{1K}^n$.[9] P1 is least likely to give in the nonexclusive game after P2 keeps (Comparisons 1 and 4). P1's giving rate in the control game is larger than this baseline, but lower than P1's giving rate in the treatment games after P2 gives (Comparisons 2 and 3). Within the treatment games, in the case where P2 gives, P1's giving rate will be similar in the exclusive and nonexclusive games (Comparison 5).

We start by considering the predictions with alternative utility specifications in which only one psychological component is considered at a time (Models A, I, and R). Under altruism alone (Model A), P1 has a positive giving probability at all five nodes, since her gift will enable P0 to have $2 rather than $0 in all cases (Comparisons 1-5 under Model A). Under reciprocity motives alone (Model R), P1 is inclined to give only after receiving a gift from P2. P1's giving rates will be greater after P2 gave in the treatment games than in the control game, even though P1 would have $3 in all cases (Comparisons 2 and 3 under Model R, $\gamma_{1G}^e > \gamma_1^c$ and $\gamma_{1G}^n > \gamma_1^c$). Under inequity aversion alone, P1 is equally likely to give at all nodes where she received $3 (in the treatment games after P2 gave or in the control game), since giving ensures that all players receive $2 (Comparisons 2, 3, and 5 under Model I).

We next consider predictions when one of the three psychological components is omitted (Models AI, IR, and AR). Without reciprocity, P1 would be equally inclined to give in the control game and after P2 gave in the treatment games (Comparisons 2, 3, and 5 under Model AI). Without altruism, P1 would never give after P2 kept in the nonexclusive game (Comparisons 1 and 4 under Model IR). Without inequity aversion (Model AR), predictions for P1's behavior are the same as in the general AIR model. Note that this means inequity aversion does not play a role in explaining Last Movers' giving behavior.

2.3.2. Predictions for Initial Movers

Rows 6–8 of Table 2 summarize predicted pairwise comparisons of giving rates for Initial Movers (P1 in the control game and P2 in the treatment games). In Model AIR, P2's giving rate is higher in the exclusive game than the nonexclusive game (Comparison 6 under Model AIR), since failure to give precludes all subsequent giving in the exclusive game. However, it is unclear whether P2 in the treatment games would be more inclined to give than P1 in the control game (Comparisons 7 and 8 under Model AIR), since altruism pushes for greater giving and inequity aversion pushes for lower giving in the treatment games.

Models A, I, and R consider the isolated role of each psychological component in predicting Initial Movers' strategies. If only altruism motivated our subjects, giving by Initial Movers would be highest in the exclusive game, since the failure to give would preclude any giving by downstream players (Comparisons 6 and 7 under Model A). If only inequity aversion motivated our subjects, P2 would be equally inclined to give in the exclusive and nonexclusive games, since she knows that failure to give would leave either P1 or P0 with nothing (Comparison 6 under Model I). With only reciprocity, P1 would only give after P2 gave and not after P2 kept in the nonexclusive game (Comparison 4 under Model R). 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 (Comparison 6 under Model R).

Models AI, IR, and AR each omit one psychological component. Without reciprocity, giving by P2 would be greater in the exclusive game than in the nonexclusive game (Comparison 6 under Model AI). P2 knows that failing to give in the exclusive game precludes P1 from giving to P0, which is undesirable since she is altruistic toward P0 and averse to inequity. However, in the nonexclusive game, the consequences of failing to give are less certain, since P1 technically can still give to P0 even if P2 did not give. Without altruism, in contrast, P2 is equally likely to give in the exclusive and nonexclusive games (Comparison 6 under Model IR). This is because P2 knows that her gift will increase P1's likelihood of giving via reciprocity motives. This means in both the treatment games, her gift generates the same likelihood of achieving equal payoffs of $2 for all players. Finally, without inequity aversion, the Initial Mover would be more inclined to give in the treatment games than in the control game, since her gift would affect more downstream players (Comparisons 7 and 8 under Model AR).

3. Experimental procedure

3.1. Implementation

Experimental sessions were implemented on Amazon Mechanical Turk (MTurk), an online platform commonly used by experimental social scientists to collect information about choices, attitudes, and opinions. The study was administered between February 23 and March 26 of 2021.[10] It has been approved by the Institutional Review Board at Michigan State University.

---

[8] We list the models' predictions for all pairwise comparisons of P1's four giving decisions, except for the comparison between γ_{1G}^e and γ_{1K}^n. The comparison of γ_{1G}^e and γ_{1K}^n is already captured by two existing comparisons. As we will discuss, we predict $\gamma_{1G}^e \sim \gamma_{1G}^n$ for all models (Comparison 5). We predict that $\gamma_{1G}^n \sim \gamma_{1K}^n$ under Models S, A, and R and that $\gamma_{1G}^n > \gamma_{1K}^n$ under all other models (Comparison 1). It follows that $\gamma_{1G}^e \sim \gamma_{1K}^n$ if the S, A, or R model is true and $\gamma_{1G}^e > \gamma_{1K}^n$ if any other model is true.

[9] Recall that the subscript $1G$ ($1K$) on γ denotes P1's giving likelihood in the case where P2 gave (kept). The superscripts $\{c, e, n\}$ denote the control, exclusive, and nonexclusive games described in Section 2.2.

[10] The experiment was conducted online since it took place during the Covid pandemic, and nonessential in-person studies were prohibited by Michigan State University.

Since our study involved three-player games, we held sessions of 9 subjects each for a total of 43 sessions.[11] A total of 408 subjects received payment for the study, but only 403 responded to all questions and were included in the full sample. All MTurk users were eligible to participate; we chose not to restrict our sample of participants based on prior performance at MTurk tasks. However, we use quality check questions to monitor subject attention and comprehension before almost all games (discussed below).

We conducted the experiment using Qualtrics software. 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 payments via MTurk within 24 hours of completion.

3.2. Experimental procedure

We used the strategy method to elicit subjects' actions at all nodes of all games. 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. Importantly, subjects 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. Subjects could not contact each other or know with whom they would be playing when they made their decisions.

All subjects proceeded through the experiment as follows. First, they were taken to the consent page, which described their rights as study participants. Next, they viewed a video that described the study and all games. 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. Throughout the session, they could click on a link that displayed the extensive-form diagram and video describing the relevant game. To check subject comprehension, we asked questions about the game rules before the exclusive and nonexclusive games. Subjects were informed they would earn an additional \$0.50 per game if they answered all questions for the game correctly on their first attempt.

All subjects played the control game first. After the control game, the order of the exclusive and nonexclusive games was randomized.[12] Each game asked subjects whether they would keep or pass their chip in each player role and each node of the extensive form game. After subjects made their giving decisions in the treatment games, we asked them about their first- and second-order beliefs regarding whether P1 would give. We discuss the details of these questions further in Online Appendix B.1, where we assess whether we can evaluate our results under dynamic reciprocity equilibrium based on subjects' reported beliefs. At the end of the study, subjects completed a demographic questionnaire. Further details of the experiment, including screenshots, are available in Appendix D.

Table 3 displays summary statistics. The top panel summarizes subject performance. Bonus payments ranged from \$0 to \$5, with a median payment of \$3 and an average payment of \$2.55. In the full sample, the average number of quality check questions answered incorrectly on the first try out of four was 2.16, with a median of 1. However, the distribution is positively skewed: Of 403 total subjects, 140 had no incorrect questions, 87 had one incorrect question, and 97 had two incorrect questions on the first try. The remaining 78 had 3 or more incorrect questions on the first try. It is likely that those who answered more than two questions incorrectly did not understand the games, so their choices may not reflect their true preferences. To exclude subjects who demonstrably struggled with the quality check questions, we define a separate subsample of subjects who answered two or more questions correctly on the first try, called the *accurate responders* sample. In robustness checks, we show that our results hold with the accurate responders sample ($N = 324$), as well as with a more limited sample of subjects who answered every question correctly on the first try ($N = 140$).

In the full sample, 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. Twenty-eight subjects took between 45 and 60 minutes. As our robustness checks will show, excluding these 30 subjects does not appreciably change the results.

Based on the demographic questionnaire, less than a third of subjects (27.5%) are women, 83% have at least an associate's degree, and 93% are employed full- or part-time. Around 68% are US citizens, 73% are US residents, and 76% are native English speakers. In terms of race and ethnicity, about half of subjects are white, 29% are Asian, 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% are 36-45 years old, 16% are 16-25 years old, and 11% of subjects are 46-65 years old. Only 1% of subjects are 65 or older.

4. Experimental results

We first assess empirical behaviors using paired one-tailed t-tests and signed-rank tests (Table 4). We then calculate the proportion of subjects whose behaviors align with model predictions (Fig. 4). Although the two methods evaluate subject behavior in different ways, they arrive at the same conclusion. Table 4 compares aggregate giving rates at different nodes, while Fig. 4 focuses on the number of subjects that adopt a given strategy. Both methods demonstrate the roles of altruism, reciprocity, and inequity aversion in explaining our experimental results.

---

[11] The exception was our first session, which was held with 30 subjects.

[12] We find no significant difference in giving rates based on game order (see Appendix Table B3).

Table 3
Summary statistics of study subjects.

Study Characteristics	
% saw exclusive game first	0.522 (0.0249)
study duration (minutes)	28.73 (0.543)
median study duration (minutes)	26.75
bonus payment	2.553 (0.0692)
median bonus payment	3
wrong answers	2.157 (0.117)
median wrong answers	1
Demographics	
% female	0.275 (0.0223)
% college graduate	0.829 (0.0188)
% employed	0.931 (0.0127)
Citizenship/residency/language fluency	
% US citizen	0.684 (0.0234)
% native English speaker	0.763 (0.0213)
% US resident	0.727 (0.0222)
Race/ethnicity	
% Black	0.129 (0.0167)
% Asian	0.293 (0.0227)
% Hispanic	0.0496 (0.0108)
% White	0.501 (0.0249))
% Other race/ethnicity	0.0273 (0.00813)
Age	
% 16-25 years old	0.159 (0.0182)
% 26-35 years old	0.496 (0.0249)
% 36-45 years old	0.223 (0.0208)
% 46-55 years old	0.0670 (0.0125)
% 56-65 years old	0.0422 (0.0100)
% 65 or older	0.0124 (0.00552)
Observations	403

Note: Summary statistics of full sample. Standard errors in parentheses.

4.1. Giving rate comparisons

In Table 4, we first compare the giving of Last Movers (P1) across games, which establishes the indirect reciprocity effect and points to the role of both altruism and reciprocity incentives in explaining pay-it-forward behavior. We then compare the actions of Initial Movers (P1 in the control game and P2 in the treatment games), which demonstrates that inequity aversion can explain P2's relatively low giving rates. In particular, we find that the knowledge that P1 could magnify P2's impact by paying P2's generosity forward does *not* increase P2's giving likelihood. Rather, P2 is reluctant to give, since giving to P1 could lead P1 to end up with more chips than her.

Rows 1–5 of Table 4 report comparisons of giving rates by Last Movers. We establish the indirect reciprocity effect by comparing P1's behavior in the treatment games after P2 gave to P1's behavior in the control game (Comparisons 2 and 3). P1's giving rate is 53.1% in the exclusive game and 54.1% in the nonexclusive game after P2 gave. Both values are significantly greater than P1's giving rate of 44.7% in the control game ($p < 0.01$ in both comparisons). 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 $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$ if she were to keep. The game design holds constant social concerns, 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 attributable to P2's kindness, rather than experimental conditions. Receiving the gift increases P1's giving likelihood by 19-21%, indicating that benefiting from another person's kindness makes subjects more likely to pay it forward. Therefore, only Models R, AR, IR, and AIR generate predictions that are consistent with subjects' behavior.

We next note that P1 has a positive probability of giving even when P2 keeps ($\hat{\gamma}_{1K}^n = 23.3\% > 0, p < 0.01$). Only altruism can explain this behavior (see Comparison 1 in Table 2), so we further rule out the R and IR models. Only the AR and AIR models remain as candidate explanations.

We then examine if the remaining predictions can differentiate between the AR and AIR models (Comparisons 4 and 5). We find that they do not, since they all predict that our findings would be supported by both the AR and the AIR models. In other words, inequity aversion does not uniquely explain Last Movers' behaviors. We therefore turn to Initial Movers' decisions.

Rows 6–8 of Table 4 reports the giving decisions of Initial Movers, which comprise of P2 in the treatment games and P1 in the control game. We begin by comparing P1's giving in the control group with P2's giving in the treatment groups (Comparisons 7 and 8). Without inequity aversion, the model would predict that P2 in the treatment games will give more than P1 in the control game. The rationale is that P2's giving should increase with the knowledge that her gift would make P1 more likely to give. However, our results show the opposite pattern. We find significantly greater giving by P1 in the control game than P2 in the treatment games

Table 4

Comparisons of giving: experimental results.

	Last Movers (P1 in all games)										
Comparison	Experimental result			Consistent with predictions?							
	giving rates	p-value from t-test	p-value from signed rank	S	A	I	R	AI	IR	AR	AIR
1: γ_{1G}^n versus γ_{1K}^n	$\hat{\gamma}_{1G}^n = 54.1\% > \hat{\gamma}_{1K}^n = 23.3\%$	$p < 0.0001$	$p < 0.0001$					✓		✓	✓
2: γ_{1G}^e versus γ_1^c	$\hat{\gamma}_{1G}^e = 53.1\% > \hat{\gamma}_1^c = 44.7\%$	$p = 0.0010$	$p = 0.0021$						✓	✓	✓
3: γ_{1G}^n versus γ_1^c	$\hat{\gamma}_{1G}^n = 54.1\% > \hat{\gamma}_1^c = 44.7\%$	$p = 0.0002$	$p = 0.0005$						✓	✓	✓
4: γ_1^c versus γ_{1K}^n	$\hat{\gamma}_1^c = 44.7\% > \hat{\gamma}_{1K}^n = 23.3\%$	$p < 0.0001$	$p < 0.0001$					✓		✓	✓
5: γ_{1G}^n versus γ_{1G}^e	$\hat{\gamma}_{1G}^n = 54.1\% \sim \hat{\gamma}_{1G}^e = 53.1\%$	$p = 0.3402$	$p = 0.6799$	✓	✓	✓	✓	✓	✓	✓	✓

	Initial Movers (P1 in the control game and P2 in the treatment games)										
Comparison	Experimental result			Consistent with predictions?							
	giving rates	p-value from t-test	p-value from signed rank	S	A	I	R	AI	IR	AR	AIR
6: γ_2^n versus γ_2^e	$\hat{\gamma}_2^e = 40.2\% \sim \hat{\gamma}_2^n = 39.0\%$	$p = 0.2542$	$p = 0.5078$			✓			✓		
7: γ_2^e versus γ_1^c	$\hat{\gamma}_2^e = 40.2\% < \hat{\gamma}_1^c = 44.7\%$	$p = 0.0156$	$p = 0.0314$			✓		✓	✓		✓
8: γ_2^n versus γ_1^c	$\hat{\gamma}_2^n = 39.0\% < \hat{\gamma}_1^c = 44.7\%$	$p = 0.0043$	$p = 0.0088$			✓		✓	✓		✓

Note: ✓ indicates that the prediction is consistent with the statistically significant experimental result.

Labels: Giving rate is denoted by γ. The superscript 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. S - standard model; A - altruism; R - reciprocity; I - inequity aversion.

(44.7% by P1 in the control game versus 40.2% and 39.0% by P2 in the exclusive and nonexclusive games respectively, $p < 0.05$). The experimental results go against the predictions of all models that do not incorporate inequity aversion.

Comparing P1 in the control group with P2 in the treatment group thus shows that including inequity aversion can align theoretical predictions with subject behaviors. Intuitively, P1 in the control game knows that by giving, she can equalize everyone's payoffs. However, P2 in the treatment games cannot equalize everyone's payoffs, since she cannot control what P1 will do after she gives. By giving, she risks ending up with less than P1. Inequity aversion would therefore push subjects to give more as P1 in the control game than P2 in the treatment games.[13]

Overall, Table 4 reports how aggregate giving rates support the AIR model. Reciprocity motives explain why P1 is more likely to give to P0 after receiving a gift from P2 (Comparisons 2 and 3). However, knowing that P1 may pay forward P2's generosity does not increase P2's chances of giving. Rather, inequity aversion explains why P2's giving rates are lower in the treatment games than P1's giving in the control game (Comparisons 7 and 8). Lastly, altruism alone can explain why P1 would give in the nonexclusive game even after P2 kept (Comparison 1).
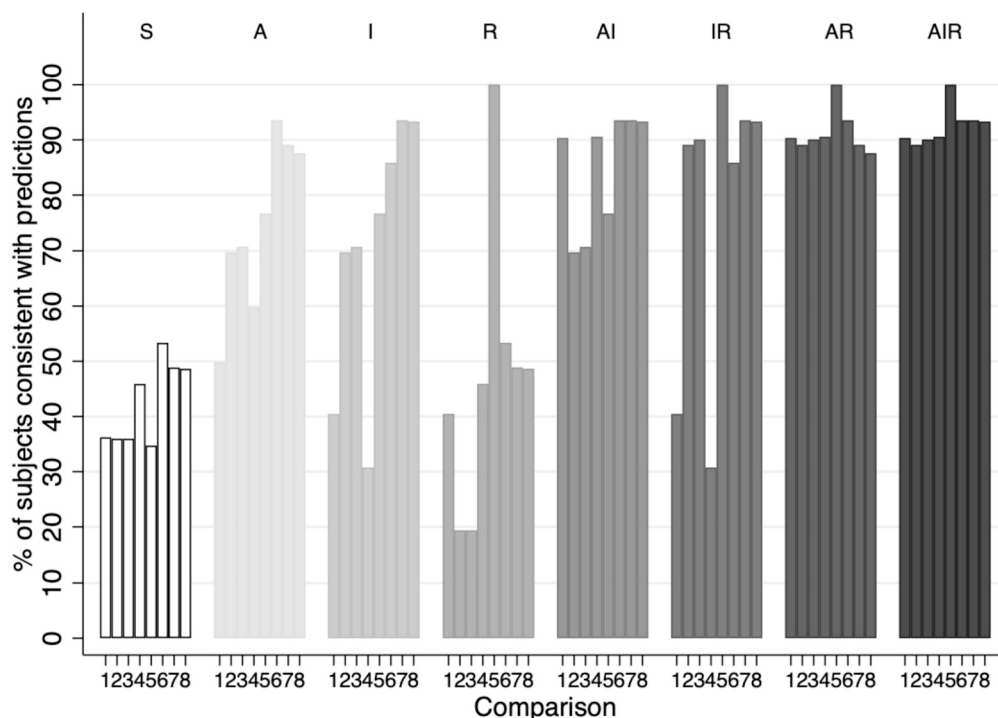
4.1.1. Robustness checks

We conduct a number of robustness checks by examining results across different samples. Across all samples, the general AIR model performed the best in aligning with theoretical predictions. Table B1 reports the results for accurate responders, who answered at least two questions correctly on the first try (panel a), subjects who only gave accurate answers (panel b), subjects who had 6 or fewer incorrect answers (panel c),[14] accurate responders who took less than 45 minutes to complete the experiment (panel d), accurate responders who saw the nonexclusive game first (panel e), and accurate responders who saw the exclusive game first (panel f). Appendix Table B3 shows that game order does not significantly impact giving rates.

4.2. Tabulating consistent strategies

In this section, we compute the proportion of subjects whose choices are consistent with each model's predictions. This alternate way of assessing model performance has two advantages over the pairwise hypothesis tests in Table 4. First, it better leverages within-subject variation by counting the number of subjects that adopt a given strategy, rather than computing aggregate giving

---

[13] Our results are also supported by the comparison of P2's giving in the exclusive and nonexclusive games (Comparison 6), which predicts greater giving by P2 in the exclusive game than the nonexclusive game. This difference is insignificant in the full sample but significant in various robustness checks (see Table B1).

[14] The number of incorrect answers differs 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.

Note: A bar represents the proportion of subjects whose experimental behavior is consistent with a model's prediction. We test each of the eight comparisons on each of the eight models.

Fig. 4. Assessing the predictive power of each model.

likelihoods at each node. Second, it allows us to quantify the importance of each psychological component in explaining empirical choices. We find that altruism is the most important, reciprocity second most important, and inequity aversion least important in explaining subject behavior.

To spell out our approach, we start with the premise that at any two decision nodes, subjects have four available strategies: (give, give), (give, keep), (keep, give), and (keep, keep). Table 2 predicts which strategies will be played given the two decision nodes that are being compared. When the prediction is ~ 0, the model predicts that giving rates at both decision nodes will be statistically indistinguishable from 0, so subjects should play (keep, keep). When the prediction is $\sim$, the model predicts equivalent actions at the two decision nodes: (give, give) or (keep, keep). Third, when the prediction is $\succ 0$, the model predicts that giving at the first node would be strictly greater than giving at the second node, which would be equivalent to 0. The only action that aligns with such a prediction is (give, keep). Lastly, when the prediction is $\succ$, the model predicts greater giving rates at the first node than the second node. This means subjects may give at both nodes, keep at both nodes, or give at the first node and keep at the second node. The only action inconsistent with the prediction of $\succ$ is (keep, give).

The exceptions to this method are Comparisons 5, 7, and 8. Under all models, Comparison 5 predicts that P1's giving inclination after P2 gave will not significantly differ between the exclusive and nonexclusive games. This does not restrict how P1's mixed strategy gets realized. Subjects who give at one node and keep at the other node do not definitively violate Comparison 5, since it is possible that they are indifferent between the two decisions and choose at random.[15] Hence, all strategies can occur even when P1's giving inclination is the same in the two games. Comparisons 7 and 8 compare P2 in the treatment games with P1 in the control game. Under any model incorporating inequity aversion, all choices are possible, depending on the specific values of the altruism and inequity aversion parameters (A, α, β).

Fig. 4 plots the proportion of subjects whose strategies align with different model predictions. Aggregate numbers are summarized in Table B2. Each comparison is listed at the bottom of the graph, and each model is listed at the top of the graph. The bars represent the proportion of subjects whose behavior is consistent with a prediction for a comparison under a given model.

Model S, which assumes that subjects only care about material payoffs, predicts that P1 would always play keep. It can only explain the behavior of the 35-36% of subjects who do so. Similarly, excluding altruism (Models R, I, and IR) fails to explain the behavior of most subjects. The model with only reciprocity (Model R) predicts that P1 must give in the treatment games if P2 gave and keep in the control game (Comparisons 2 and 3). In our data, only 19% of subjects exhibit this behavior, since 81% of subjects give in the control game or keep in the treatment games after P2 gave. Models I and IR predict that P1s must give in the control game

---

[15] This result speaks to the necessity of incorporating mixed strategies to account for indifference. Only considering pure strategies would lead us to infer that a subject's behavior reflected their dominant strategy, and that this strategy was strictly preferred above other alternatives.

and keep in the nonexclusive game after P2 kept (Comparison 4). They fail to explain the behaviors of the 61% of P1s who keep in the control game or give in the nonexclusive game after P2 kept. Together, these results establish the importance of altruism in our "pay it forward" games.

Excluding reciprocity motives would also fail to explain the behavior of the majority of subjects. Since we have ruled out Model I, we focus on Models A and AI. Model A, which only allows for altruism, predicts that P1 will be equally inclined to give independent of whether P2 gave or kept in the nonexclusive game (Comparison 1). It cannot explain behavior for the 50% of subjects whose decisions as P1 differ based on whether P2 gave or kept. Model AI predicts that, absent reciprocity motives, P1 should have equal giving inclinations in the control game and the treatment games after P2 gave (Comparisons 2 and 3). It cannot explain the behavior of the 30% of subjects who choose different actions at these nodes.

We are then left with the model with altruism and reciprocity (Model AR) and the model with altruism, reciprocity, and inequity aversion (Model AIR). The two models generate identical predictions for P1's behavior (Comparisons 1-5), but the AIR model performs slightly better in explaining 93% of P2's behavior (Comparisons 6-8). The AR model explains 88-89% of P2's behavior. It predicts that in the absence of inequity aversion, P2's giving in the treatment games should be greater than P1's giving in the control game (Comparisons 7 and 8). These predictions are at odds with the t-test results from Table 4, where aggregate giving rates are higher for P1 in the control game than P2 in the treatment games. They cannot explain the 10-12% of subjects that keep as P2 in the treatment games but give as P1 in the control game. The AIR model better rationalizes the behavior of these subjects, since inequity aversion can explain why they give in the control game but not in the treatment games provided individual parameters (A, Z, α, β).

Finally, we evaluate the relative importance of altruism, reciprocity, and inequity aversion in explaining the experimental data. Taking into account all propositions, models without altruism can only explain 19-35% of subjects' behavior. Models without reciprocity can explain 31-70% of subjects' behavior, and models without inequity aversion can explain 19-88% of subjects' behavior. By comparison, the AIR model can explain behavior for 90% of subjects. This means that adding altruism to the IR model increases the proportion of subjects whose behavior can be explained by $89.1-30.8 = 58.3\%$, or 235 subjects. Adding reciprocity to the AI model increases explanatory power from 69.7% to 89.1% of subjects, a gain of 19.4% or 78 subjects. Adding inequity aversion to the AR model increases explanatory power from 87.6% to 89.1%, a gain of 1.5% or 6 subjects. Altruism and reciprocity substantially increase model performance, while there is only a marginal improvement from incorporating inequity aversion.

4.3. Comparing the two empirical approaches

Our model predicts differences in giving inclinations for the same subject across different game nodes. We evaluate model predictions with experimental data using two distinct approaches. In the first approach, pairwise hypothesis tests determine whether aggregate differences in giving rates across game nodes are statistically significant (Table 4). We find that the results can only be supported by the model that incorporates altruism, reciprocity, and inequity aversion (AIR). The second approach quantifies the importance of altruism, reciprocity, and inequity aversion by tabulating the proportion of subjects whose strategy follows predicted patterns across a given pair of game nodes (Fig. 4). We find that omitting altruism or reciprocity would sacrifice considerable explanatory power. However, inequity aversion plays a marginal role.

Despite substantive differences in approach, both methodologies arrive at the same conclusions. They show that different psychological components explain behavior for different players. Altruism and reciprocity are key for describing P1's behavior, and therefore explain why receiving help might lead one to help an unrelated third party. Inequity aversion plays no role, as the AIR model performs no better than the AR model in explaining P1's behavior. Rather, inequity aversion helps explain P2's behavior. P2 is less likely to give in the treatment games than P1 in the control game, since in the latter case P1 can equalize payoffs across all players. In contrast, by giving to P1, P2 risks the possibility that P1 will keep and end up with greater final payoffs than P2.

These three psychological components account for behavior in different ways. For the manager seeking to promote helping behavior in the workplace, our results suggest that appealing to reciprocity and altruism will mainly affect how people pay forward help they have received in the past—how chains of generosity continue after they are launched. Meanwhile, inequity aversion may impede launching the chain of kindness in the first place. A supervisor considering whether to mentor one subordinate over others may be concerned about exhibiting favoritism, and therefore mentor no one. This could then lower the likelihood that her subordinates "pay forward" mentoring in future years, after they have become supervisors themselves.

4.4. Additional considerations

Our definition of dynamic reciprocity equilibrium assumes consistency: subjects have correct beliefs regarding other players' behavior strategies. We elicit first- and second-order beliefs of other players' behaviors throughout our experiment. Across subjects, beliefs are close to true behaviors, which is a necessary condition for experimentally testing our model predictions within our concept of dynamic reciprocity equilibrium. For details on the elicitation and results, see Appendix B.1.

While additional tests of the consistency assumption are beyond the scope of this paper, we note that our empirical analysis does not rely on it. This is because our approach involves comparing giving inclinations across decision nodes. More precisely, if errors in beliefs are subject-specific, comparing across decision nodes would difference them out. For example, if a subject systematically underestimates others' giving inclinations, that error in belief should be differenced out when comparing her giving as P2 in the exclusive game versus the nonexclusive game.

Second, we note that inequity aversion plays a marginal role in explaining subject behaviors. We develop and test an alternative explanation behind our findings: that giving decisions are motivated by how much credit one can receive for improving others'

payoffs. Our alternate model formulates explicit predictions on the relationship between giving, credit, and beliefs. For each game, we ask subjects about the credit each player should receive for the payoff of the other two players. Finally, we test our model predictions against our experimental results. Overall, we do not find sufficient evidence that credit for others' payoffs motivates giving. Appendix C discusses relevant theoretical and experimental results.

5. Conclusion

We evaluate the importance of indirect reciprocity, altruism, and inequity aversion in motivating pay-it-forward behavior. We establish a psychological game-theoretic framework that formulates predictions for giving behavior under different models of prosocial behavior. We then test these predictions using a novel experiment that demonstrates the existence of indirect reciprocal exchange while controlling for alternative explanations such as income effects, relative payoffs, and social image considerations. The combination of experimental design and theoretical framework enables us to exploit within-subject variation when comparing across various game nodes. That is, by assuming that subjects have constant prosocial preferences across games, we isolate distinct patterns in giving behavior in different games and player roles. We find that indirect reciprocity incentives are critical to explain the pay-it-forward behavior, where receiving a gift makes P1 more likely to give. However, knowledge that P1 may pay forward P2's generosity does not appear to encourage giving by P2, even though P2's generosity would have been magnified by P1's pay-it-forward behavior. Rather, inequity aversion provides one explanation as to why P2 is less likely to give, which makes the transmission of generosity unlikely to start in the first place.

Our findings address the question of how generosity spreads within communities. This phenomenon has been documented by prior work, but we know little about the conditions that start and maintain this spread. We provide experimental evidence that people pay forward kind acts to unrelated others; namely, that kindness engenders further kindness. However, the knowledge that others may pay forward your kindness does not make you more likely to help. Chains of generosity easily continue once started, but are relatively difficult to start. These results speak to the stability of culture across different contexts. In a workplace where a new employee is mentored and helped, she is likely to help other newcomers in the future. In contrast, if a new employee was left to fend for herself, she may not think to help newcomers in the future even when she could. The tendency of subjects to reciprocate help with help (and its absence with no help) can contribute to the formation of social norms and behavioral conduct within organizations, neighborhoods, and other social settings.

Declaration of competing interest

The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Appendix 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 and consistency with Dufwenberg and Kirchsteiger (2004), 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

$$\Sigma_i^e := \Bigg\{ \sigma_i \in \Sigma_i \Big| \nexists \hat{\sigma}_i \in \Sigma_i \text{ such that}$$

$$\Big(\forall (h \in H, \sigma_{-i} \in \Sigma_{-i}, k \in N), \pi_k^h \left(\hat{\sigma}_i^{[h]}, \sigma_{-i}^{[h]} \right) \geq \pi_k^h \left(\sigma_i^{[h]}, \sigma_{-i}^{[h]} \right) \Big) \wedge$$

$$\Big(\exists (h \in H, \sigma_{-i} \in \Sigma_{-i}, k \in N) \text{ such that } \pi_k^h \left(\hat{\sigma}_i^{[h]}, \sigma_{-i}^{[h]} \right) > \pi_k^h \left(\sigma_i^{[h]}, \sigma_{-i}^{[h]} \right) \Big) \Bigg\}.$$

A.2. Proof of the theorem on equilibrium existence

Proof of Theorem 1. Let $\Delta(C_i(h))$ denote player i's set of (potentially random) choices available at history $h \in H$. For any choice $c_i \in \Delta(C_i(h))$, let $\sigma_i(h, c_i)$ denote player i's strategy that specifies the choice c_i at h, but is the same as $\sigma_i^{[h]}$ otherwise—i.e., at every history in $H \setminus \{h\}$. Let Σ_i denote i's set of behavior strategies and define $\Sigma = \prod_{i \in N} \Sigma_i$. Define correspondence $B_{i,h} : \Sigma \rightarrow \Delta(C_i(h))$ by

$$B_{i,h}(\sigma) = \underset{x \in \Delta(C_i(h))}{\arg\max} \, u_i^h \left(\sigma_i(h, x), (\sigma_j^{[h]}, (\sigma_k^{[h]})_{k \neq j})_{j \neq i} \right),$$

and define correspondence $B : \Sigma \rightarrow \Pi_{(i,h) \in N \times H} \Delta(C_i(h))$ by

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

The set $\prod_{(i,h)\in N\times H} B_{i,h}(\sigma)$ is topologically equivalent to the set Σ, so $B : \Sigma \to \Pi_{(i,h)\in N\times H}\Delta(C_i(h))$ is equivalent to a correspondence $\gamma : \Sigma \to \Sigma$ (which is a direct redefinition of B). Every fixed point of γ is an equilibrium. To see this, note that a fixed point $B_{i,h}$ satisfies utility maximization under consistent beliefs. Here, because $B_{i,h}$ specifies the optimal choices at each $h \in H$, altogether, $B_{i,h}$ specifies the optimal strategies in $\Sigma_i(h,x)$. Hence, B 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 $B_{i,h}$ is nonempty, closed-valued, and upper hemicontinuous, since $\Delta(C_i(h))$ is nonempty and compact and u_i is continuous (since π_i, κ_{ij}, and λ_{ijk} are all continuous). In addition, $B_{i,h}$ is convex-valued, since $\Delta(C_i(h))$ is convex and u_i is linear—and hence quasiconcave—in i's own choice (because u_i is linear in π_i, which is linear in σ_i). Hence, $B_{i,h}$ is nonempty, closed-valued, upper hemicontinuous, and convex-valued. These properties extend to B and γ. It follows by Kakutani's fixed-point theorem that γ admits a fixed point. $\quad\square$

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 $u_1(g_1) = 2 + 4A_1$. P1's utility from keeping is $u_1(k_1) = 3 + 2A_1 - \beta_1(3 - 2)/2 - \beta_1(3 - 0)/2 = 3 + 2A_1 - 2\beta_1$. P1 prefers giving if and only if $u_1(g_1) = 2 + 4A_1 \geq u_1(k_1) = 3 + 2A_1 - 2\beta_1$, that is, $2A_1 + 2\beta_1 \geq 1$. $\quad\square$

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 $u_1(g_{1G}, \gamma''_{1G}) = 2 + 4A_1 + Z_1(+1)(1 - \gamma''_{1G}/2)$, and P1's utility from keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$, is $u_1(k_{1G}, \gamma''_{1G}) = 3 + 2A_1 - 2\beta_1 + Z_1(-1)(1 - \gamma''_{1G}/2)$. Therefore, P1's utility from giving with probability γ_{1G} is $u_1(\gamma_{1G}, \gamma''_{1G}) = \gamma_{1G}[-1 + 2A_1 + 2\beta_1 + Z_1(2 - \gamma''_{1G})] + 3 + 2A_1 - 2\beta_1 - Z_1(1 - \gamma''_{1G}/2)$. If $2A_1 + 2\beta_1 + Z_1 \geq 1$, then $\gamma_{1G} = 1$. If $2A_1 + 2\beta_1 + 2Z_1 \leq 1$, then $\gamma_{1G} = 0$. If $2A_1 + 2\beta_1 + Z_1 < 1 < 2A_1 + 2\beta_1 + 2Z_1$, then $-1 + 2A_1 + 2\beta_1 + Z_1(2 - \gamma_{1G}) = 0$, which rearranges to $\gamma_{1G} = 2 - \frac{1 - 2A_1 - 2\beta_1}{Z_1}$. Therefore, in equilibrium, $\gamma^e_{1G} = [\![2 + \frac{2A_1 + 2\beta_1 - 1}{Z_1}]\!]$.

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 $\gamma'_{1G}(2 + 4A_2) + (1 - \gamma'_{1G})(2 + 3A_2 - \alpha_2/2 - \beta_2) = 2 + 4A_2 - (1 - \gamma'_{1G})(A_2 + \alpha_2/2 + \beta_2)$. P2's utility from keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (3, 1, 0)$, is $3 + 1A_2 - \beta_2(3 - 1)/2 - \beta_2(3 - 0)/2 = 3 + A_2 - 5\beta_2/2$. P2 prefers giving if $2 + 4A_2 - (1 - \gamma'_{1G})(A_2 + \alpha_2/2 + \beta_2) \geq 3 + A_2 - 5\beta_2/2$, which is simplified to $3A_2 + 5\beta_2/2 - (1 - \gamma'_{1G})(A_2 + \alpha_2/2 + \beta_2) \geq 1$. In equilibrium, $\gamma'_{1G} = \gamma_{1G}$, so the inequality is rearranged to $(2 + \gamma_{1G})A_2 - (1 - \gamma_{1G})\alpha_2/2 + (3/2 + \gamma_{1G})\beta_2 \geq 1$. $\quad\square$

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

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 $u_1(g_2, g_{1G}, \cdots) = 2 + 4A_1 + Z_1(1 - \gamma''_{1G}/2 + \gamma''_{1K}/2)$. P1's utility of keeping, which yields material payoffs $(\pi_2, \pi_1, \pi_0) = (2, 3, 0)$, is $u_1(g_2, k_{1G}, \cdots) = 3 + 2A_1 - \beta_1(3 - 2)/2 - \beta_1(3 - 0)/2 - Z_1(1 - \gamma''_{1G}/2 + \gamma''_{1K}/2) = 3 + 2A_1 - 2\beta_1 - Z_1(1 - \gamma''_{1G}/2 + \gamma''_{1K}/2)$. Hence, $u_1(g_2, g_{1G}, \cdots) - u_1(g_2, k_{1G}, \cdots) = -1 + 2A_1 + 2\beta_1 + Z_1(2 - \gamma''_{1G} + \gamma''_{1K})$.

Comparing the net benefit of giving after P2 gave and that after P2 kept, 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 kept, she will also choose to give after P2 gave. In other words, P1 is more inclined to give after P2 gave than after P2 kept: $\gamma_{1G} \geq \gamma_{1K}$. 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 $2A_1 - 3\alpha_1/2 - \beta_1/2 - 2Z_1 \geq 1$.
2. Strategies $\gamma_{1G} = 1$ and $0 < \gamma_{1K} < 1$ are supported in equilibrium when and only when $2A_1 - 3\alpha_1/2 - \beta_1/2 - 2Z_1 \leq 1 \leq 2A_1 - 3\alpha_1/2 - \beta_1/2 - Z_1$. In this case, $2A_1 - 3\alpha_1/2 - \beta_1/2 - Z_1(1 + \gamma_{1K}) = 1$, which is rearranged to $\gamma_{1K} = \frac{(2A_1 - 3\alpha_1/2 - \beta_1/2 - 1 - Z_1)}{Z_1}$.
3. Strategies $\gamma_{1G} = 1$ and $\gamma_{1K} = 0$ are supported in equilibrium when and only when $2A_1 - 3\alpha_1/2 - \beta_1/2 - Z_1 \leq 1 \leq 2A_1 + 2\beta_1 + Z_1$.
4. Strategies $0 < \gamma_{1G} < 1$ and $\gamma_{1K} = 0$ are supported in equilibrium when and only when $2A_1 + 2\beta_1 + Z_1 \leq 1 \leq 2A_1 + 2\beta_1 + 2Z_1$. In this case, $2A_1 + 2\beta_1 + Z_1(2 - \gamma_{1G}) = 1$, which is rearranged to $\gamma_{1G} = 2 + \frac{(2A_1 + 2\beta_1 - 1)}{Z_1}$.
5. Strategies $\gamma_{1G} = 0$ and $\gamma_{1K} = 0$ are supported in equilibrium when and only when $2A_1 + 2\beta_1 + 2Z_1 \leq 1$.

In summary, in the nonexclusive game, P1 gives with probability $\gamma^n_{1G} = [\![\frac{(2A_1 + 2\beta_1 - 1)}{Z_1} + 2]\!]$ after P2 gave, and gives with probability $\gamma^n_{1K} = [\![\frac{(2A_1 - 3\alpha_1/2 - \beta_1/2 - 1)}{Z_1} - 1]\!]$ after P2 kept.

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 $(2 + 4A_2)\gamma'_{1G} + (2 + 3A_2 - \alpha_2/2 - \beta_2)(1 - \gamma'_{1G}) = 2 + (3 + \gamma'_{1G})A_2 - (1 - \gamma'_{1G})(\alpha_2/2 + \beta_2)$. 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 $\gamma'_{1K}[3 + 2A_2 - \beta_2(3-0)/2 - \beta_2(3-2)/2] + (1 - \gamma'_{1K})[3 + A_2 - \beta_2(3-0)/2 - \beta_2(3-1)/2] = 3 + (1 + \gamma'_{1K})A_2 - (5 - \gamma'_{1K})\beta_2/2$. P2 prefers giving if and only if $2 + (3 + \gamma'_{1G})A_2 - (1 - \gamma'_{1G})(\alpha_2/2 + \beta_2) \geq 3 + (1 + \gamma'_{1K})A_2 - (5 - \gamma'_{1K})\beta_2/2$, which, as $\gamma_{1G} = \gamma'_{1G}$ and $\gamma_{1K} = \gamma'_{1K}$ in equilibrium, is rearranged to $(2 + \gamma_{1G} - \gamma_{1K})A_2 - (1/2 - \gamma_{1G}/2)\alpha_2 + (3/2 + \gamma_{1G} - \gamma_{1K}/2)\beta_2 \geq 1$. □

A.4. Comparisons of giving: theoretical predictions

In this section, we formulate propositions for the full model that incorporates altruism, inequity aversion, and reciprocity (Model AIR). The predictions for all other models can be derived from setting appropriate factors to zero.

A.4.1. Predictions for Last Movers

P1 is the Last Mover in all three games. Fig. 2 shows P1's equilibrium giving rates for different altruistic factors. The comparisons are unambiguous if giving is motivated by altruism and reciprocity: $\gamma^e_{1G} \sim \gamma^n_{1G} > \gamma^c_1 > \gamma^n_{1K}$. There are $4 \times 3/2 = 6$ different pairwise comparisons for the four decisions. We do not directly compare γ^e_{1G} versus γ^n_{1K}, since the results follow transitively from comparing γ^e_{1G} versus γ^n_{1G} and γ^n_{1G} versus γ^n_{1K}. We discuss the remaining five pairwise comparisons in the following five propositions.

First, compare P1's two giving decisions in the nonexclusive game, after P2 gave versus after P2 kept. Regardless of P2's choice, P1 incurs the same material loss (1 unit) and altruistic gain ($2A_1$ units) from giving. However, if P1 chooses to give after P2 kept, she incurs a larger inequity aversion loss and reciprocity loss than if she chooses to give after P2 gave.

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

As argued in Section 2.2, the equilibrium giving probabilities cannot be strictly between 0 and 1 at both information sets: If P1 is indifferent at one node, she is not indifferent at the other. Proposition 1 implies that when P1 chooses to give after P2 kept, P1 must also give after P2 gave. When P1 chooses to keep after P2 gave, P1 must also keep after P2 kept. It is possible that P1 gives after P2 gave and keeps after P2 kept, but never possible for P1 to keep after P2 gave and give after P2 kept.

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

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

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

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

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

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

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

Proposition 5. *P1 is equally inclined to give after P2 gave 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 4.2.

A.4.2. Predictions for Initial Movers

We next compare the Initial Movers in these games: P1 in the control game and P2 in the treatment games. To summarize, we predict that the Initial Mover's giving rate is higher in the exclusive game than in the nonexclusive game (Fig. 3), 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, P2's incentives to give are greater in the exclusive than the nonexclusive 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 kept in the nonexclusive game will increase P2's expected utility from keeping by $\gamma_{1K}^n A_2$ from altruism and $\gamma_{1K}^n \beta_2/2$ from having more equal payoffs. Fig. 3 depicts the comparison of giving rates for initial movers.

Proposition 6. *P2 in the exclusive game is more inclined to give than P2 in the nonexclusive game.*

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 $Z_1 = Z_2 \equiv Z$. Here, altruism and inequity aversion might work in opposite directions. By giving, P1 in the control game gets an altruistic payoff of $2A$, and P2 in the exclusive game gets an altruistic payoff of $(2 + \gamma_{1G}^e)A$, because P1 in the exclusive game generates additional $\gamma_{1G}^e A$ units of altruistic payoff for P2 by passing to P0. Therefore, altruism increases P2's inclination to give in the exclusive game.

However, inequity aversion would create the opposite effect. If P2 gives, P1 might keep and end up with higher final payoffs than her. She would then suffer disutility $(3 - 2)\alpha/2 = \alpha/2$ 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 expected loss is $(1 - \gamma_{1G}^e)\alpha/2$. Furthermore, P2 does not have a sure chance of equalizing payoffs in the exclusive game since P1 may choose to keep after P2 gave, while in the control game P1 will certainly equalize payoffs by giving.

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

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

A.4.3. Proofs of propositions on equilibrium giving comparisons

Proof of Proposition 1. In the nonexclusive game, P1 gives with probability $\gamma_{1G}^n = \llbracket \frac{(2A_1 + 2\beta_1 - 1)}{Z_1} + 2 \rrbracket$ after P2 gave, and P1 gives with probability $\gamma_{1K}^n = \llbracket \frac{(2A_1 - 3\alpha_1/2 - \beta_1/2 - 1)}{Z_1} - 1 \rrbracket$ after P2 kept. Since $(2A_1 + 2\beta_1 - 1)Z_1 + 2 > \frac{(2A_1 - 3\alpha_1/2 - \beta_1/2 - 1)}{Z_1} - 1$ for any combination of nonnegative parameters A_1, α_1, β_1, and Z_1, $\gamma_{1G}^n \geq \gamma_{1K}^n$, and the inequality is strict as long as $Z_1 \neq 0$. $\square$

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

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

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

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

When $Z_1 = 0$, the condition for $\gamma_{1G}^e = 1$ and the condition for $\gamma_1^c = 1$ coincide, and the condition for $\gamma_{1G}^e = 0$ and the condition for $\gamma_1^c = 0$ also coincide. When $Z_1 > 0$, the set of parameters for $\gamma_{1G}^e = 1$ is a strict superset of that for $\gamma_1^c = 1$, and the set of parameters for $\gamma_{1G}^e = 0$ is a strict subset of that for $\gamma_1^c = 0$. For the set range of parameters for $0 < \gamma_{1G}^e < 1$, $\gamma_1^c = 0$. Hence, $\gamma_{1G}^e \geq \gamma_1^c$ for any combination of parameters. Hence, $\gamma_{1G}^e \succ \gamma_1^c$ $\square$

Proof of Proposition 3. The proof mimics that of Proposition 2, with superscripts e replaced by superscripts n. Alternatively, by Proposition 5, $\gamma_{1G}^e \sim \gamma_{1G}^n$, so by transitivity of the inclination, $\gamma_{1G}^n \succ \gamma_1^c$. $\square$

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

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

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

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

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, $\gamma_1^c > \gamma_{1K}^n$. $\square$

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

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 $B^e \equiv (2 + \gamma_{1G}^e)A_2 - (1/2 - \gamma_{1G}^e/2)\alpha_2 + (3/2 + \gamma_{1G}^e)\beta_2 - 1$. 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 $B^n \equiv (2 + \gamma_{1G}^n - \gamma_{1K}^n)A_2 - (1/2 - \gamma_{1G}^n/2)\alpha_2 + (3/2 + \gamma_{1G}^n - \gamma_{1K}^n)\beta_2 - 1$. By Proposition 5, $\gamma_{1K}^n = \gamma_{1G}^n$. Then, $B^e - B^n = \gamma_{1K}^n A_2 + \gamma_{1K}^n \beta_2$. 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. $\square$

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 $B^c \equiv 2A_1 + 2\beta_1 - 1$. 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 $B^e \equiv (2 + \gamma_{1G}^e)A_2 - (1/2 - \gamma_{1G}^e/2)\alpha_2 + (3/2 + \gamma_{1G}^e)\beta_2 - 1$. 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 $B^e - B^c = \gamma_{1G}^e A - (1 - \gamma_{1G}^e)\alpha/2 + (\gamma_{1G}^e - 1/2)\beta$. 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, $\gamma_{1G}^e A + (\gamma_{1G}^e - 1/2)\beta \leq (1 - \gamma_{1G}^e)\alpha/2$. $\square$

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 $B^c \equiv 2A_1 + 2\beta_1 - 1$. 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 $B^n \equiv (2 + \gamma_{1G}^n - \gamma_{1K}^n)A_2 - (1/2 - \gamma_{1G}^n/2)\alpha_2 + (3/2 + \gamma_{1G}^n - \gamma_{1K}^n)\beta_2 - 1$. 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 $B^n - B^c = (\gamma_{1G}^n - \gamma_{1K}^n)A - (1/2 - \gamma_{1G}^n/2)\alpha + (\gamma_{1G}^n - \gamma_{1K}^n - 1/2)\beta$. 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, $(\gamma_{1G}^n - \gamma_{1K}^n)A + (\gamma_{1G}^n - \gamma_{1K}^n - 1/2)\beta \leq (1/2 - \gamma_{1G}^n/2)\alpha$. $\square$

Appendix. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2025.07.002.

Data availability

Data will be made available on request.

References

Attanasi, G., Rimbaud, C., Villeval, M.C., 2019. Embezzlement and guilt aversion. J. Econ. Behav. Organ. 167, 409–429.
Bartlett, M.Y., DeSteno, D., 2006. Gratitude and prosocial behavior: helping when it costs you. Psychol. Sci. 17 (4), 319–325.
Battigalli, P., Catonini, E., De Vito, N., 2025. Game theory: analysis of strategic thinking. Manuscript.
Battigalli, P., Corrao, R., Dufwenberg, M., 2019. Incorporating belief-dependent motivation in games. J. Econ. Behav. Organ. 167, 185–218.
Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. J. Econ. Theory 144 (1), 1–35.
Battigalli, P., Dufwenberg, M., 2022. Belief-dependent motivations and psychological game theory. J. Econ. Lit. 60 (3), 833–882.
Ben-Ner, A., Putterman, L., Kong, F., Magan, D., 2004. Reciprocity in a two-part dictator game. J. Econ. Behav. Organ. 53 (3), 333–352.
Berger, U., 2011. Learning to cooperate via indirect reciprocity. Games Econ. Behav. 72 (1), 30–37.
Bolton, G.E., Katok, E., Ockenfels, A., 2005. Cooperation among strangers with limited information about reputation. J. Public Econ. 89 (8), 1457–1468.
Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. Am. Econ. Rev. 90 (1), 166–193.
Charness, G., Du, N., Yang, C.-L., 2011. Trust and trustworthiness reputations in an investment game. Games Econ. Behav. 72 (2), 361–375.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Q. J. Econ. 117 (3), 817–869.
Cox, J.C., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. Games Econ. Behav. 59 (1), 17–45.
Cox, J.C., Friedman, D., Sadiraj, V., 2008. Revealed altruism. Econometrica 76 (1), 31–69.
Desteno, D., Bartlett, M., Wormwood, J., Williams, L., Dickens, L., 2010. Gratitude as moral sentiment: emotion-guided cooperation in economic exchange. Emotion 10, 289–293.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games Econ. Behav. 47 (2), 268–298.
Ebrahimji, A., 2020. Over 900 cars paid for each other's meals at a Dairy Queen drive-thru in Minnesota [online]. Posted 09-December-2020.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games Econ. Behav. 54 (2), 293–315.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114 (3), 817–868.
Gaudeul, A., Keser, C., Müller, S., 2021. The evolution of morals under indirect reciprocity. Games Econ. Behav. 126, 251–277.
Gong, B., Yang, C.-L., 2019. Cooperation through indirect reciprocity: the impact of higher-order history. Games Econ. Behav. 118, 316–341.
Gray, K., Ward, A.F., Norton, M.I., 2014. Paying it forward: generalized reciprocity and the limits of generosity. J. Exp. Psychol. Gen. 143 (1), 247.
Gul, F., Pesendorfer, W., 2016. Interdependent preference models as a theory of intentions. J. Econ. Theory 165, 179–208.
Heller, Y., Mohlin, E., 2017. Observations on cooperation. Rev. Econ. Stud. 85 (4), 2253–2282.
Herne, K., Lappalainen, O., Kestilä-Kekkonen, E., 2013. Experimental comparison of direct, general, and indirect reciprocity. J. Socio-Econ. 45, 38–46.

Hu, Y., Ma, J., Luan, Z., Dubas, J.S., Xi, J., 2019. Adolescent indirect reciprocity: evidence from incentivized economic paradigms. J. Adolesc. 74, 221–228.

Iwagami, A., Masuda, N., 2010. Upstream reciprocity in heterogeneous networks. J. Theor. Biol. 265 (3), 297–305.

Jiang, L., Wu, J., 2019. Belief-updating rule and sequential reciprocity. Games Econ. Behav. 113, 770–780.

Kuhn, H.W., 1953. Extensive Games and the Problem of Information. Princeton University Press, p. 193.

Levine, D.K., 1998. Modeling altruism and spitefulness in experiments. Rev. Econ. Dyn. 1 (3), 593–622.

Malmendier, U., te Velde, V.L., Weber, R.A., 2014. Rethinking reciprocity. Annu. Rev. Econ. 6 (1), 849–874.

Melamed, D., Simpson, B., Abernathy, J., 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., Leibbrandt, A., 2018. Indirect reciprocity and prosocial behaviour: evidence from a natural field experiment. Econ. J. 128 (611), 1683–1699.

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

Nowak, M.A., Sigmund, K., 1998a. The dynamics of indirect reciprocity. J. Theor. Biol. 194 (4), 561–574.

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

Ohtsuki, H., Iwasa, Y., 2006. The leading eight: social norms that can maintain cooperation by indirect reciprocity. J. Theor. Biol. 239 (4), 435–444.

Rabin, M., 1993. Incorporating fairness into game theory and economics. Am. Econ. Rev. 83 (5), 1281–1302.

Roth, A.E., Sonmez, T., Unver, M.U., 2004. Kidney exchange. Q. J. Econ. 119 (2), 457–488.

Seinen, I., Schram, A., 2006. Social status and group norms: indirect reciprocity in a repeated helping experiment. Eur. Econ. Rev. 50 (3), 581–602.

Simpson, B., Harrell, A., Melamed, D., Heiserman, N., Negraia, D.V., 2018. The roots of reciprocity: gratitude and reputation in generalized exchange systems. Am. Sociol. Rev. 83 (1), 88–110.

Sobel, J., 2005. Interdependent preferences and reciprocity. J. Econ. Lit. 43 (2), 392–436.

Tsvetkova, M., Macy, M.W., 2014. The social contagion of generosity. PLoS ONE 9 (2), e87275.

van Apeldoorn, J., Schram, A., 2016. Indirect reciprocity: a field experiment. PLoS ONE 11 (4), e0152076.

Wu, J., 2018. Indirect higher order beliefs and cooperation. Exp. Econ. 21 (4), 858–876.

Zeckhauser, R., Swanson, J., Lockwood, K., 2006. The value of reputation on eBay: a controlled experiment. Exp. Econ. 9, 79–101.