Behavioral Visibility Increases Altruistic Preferences and Their Social Norm Mechanisms

Authors: Huang Xinru, Li Jian, Ni Yinmei, Li Jian, Ni Yinmei

Date: 2022-09-09T00:00:00+00:00

Abstract

Across different types of social preferences, researchers have devoted considerable attention to altruistic preferences and their signaling function. This study investigates how decision-makers' altruistic preferences in dictator games are influenced by the visibility of allocation schemes to recipients. Experiment 1, employing a combination of behavioral experiments and computational modeling, found that regardless of whether in choice or rating conditions, allocators exhibited a greater degree of altruistic preference when their behavior was visible to recipients compared to when it was invisible. Furthermore, compared to rating conditions, individuals placed greater emphasis on allocation efficiency when making choices. Experiment 2 used only choice conditions and manipulated social norms, finding that the effect of behavior visibility in increasing altruistic preferences depends on altruistic social norms; when non-altruistic social norms were present, the influence of behavior visibility diminished. The results of this study indicate that under altruistic social norms, when behavior is visible to recipients, people exhibit greater altruistic preferences.

Full Text

Behavioral Visibility Increases Altruistic Preference and Its Social Norm Mechanism

HUANG Xinru[1], **LI Jian**[1,2], **NI Yinmei**[1] [1]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100871, China [2]IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871, China

Abstract

Background: In social economic decisions, individuals' behavior depends not only on their own potential payoffs but also on those of others. Social preference refers to the degree and direction of considering others' payoffs relative to self-interest (Glimcher & Fehr, 2013). Among various social preference types, researchers have paid considerable attention to altruistic preference and its signaling function. This study investigates how the visibility of allocators' behavior to recipients influences altruistic preference in a dictator game.

Methods: Experiment 1 employed a 2 (visibility: visible vs. invisible) × 2 (response type: choice vs. rating) × 2 (inequity type: advantageous vs. disadvantageous) within-subjects design. Participants (N = 38) completed a modified dictator game where they either made allocation choices or rated given allocations. Their behavior was either visible to their partner or remained private. Experiment 2 used a 2 (visibility: visible vs. invisible) × 2 (social norm: altruistic vs. non-altruistic) within-subjects design (N = 53) to examine how social norms modulate the visibility effect. Social norms were manipulated by presenting the purported choice proportions of previous participants.

Results: Experiment 1 revealed that behavioral visibility increased altruistic preference across both choice and rating conditions. Specifically, when behavior was visible, participants showed greater advantageous inequity aversion (AIA) and reduced disadvantageous inequity aversion (DIA), indicating higher altruism. Computational modeling using the Fehr-Schmidt inequity aversion model confirmed these effects. Additionally, participants showed greater concern for allocation efficiency in the choice condition compared to the rating condition. Experiment 2 demonstrated that the visibility effect on altruistic preference depended on social norms: under altruistic norms, visibility significantly increased altruistic preference, whereas this effect diminished under non-altruistic norms.

Conclusion: The findings indicate that when behavior is visible to recipients, individuals exhibit greater altruistic preference, particularly under altruistic social norms. This supports the signaling function of altruistic behavior and reveals the crucial role of social norms in moderating the relationship between behavioral visibility and altruistic preference.

Keywords: social preference, altruism, behavioral visibility, response type, social norm

1 Introduction

In economic decisions involving interpersonal interaction, individuals' behavior often depends not only on their own potential payoffs but also on those of others. Social preference refers to the degree and direction of considering others' payoffs relative to self-interest (Glimcher & Fehr, 2013). Social preferences encompass multiple types. For instance, social value orientation theory classifies social preferences into altruistic, prosocial, individualistic, and competitive categories, with prosocial preferences further divided into those pursuing total payoff maximization and those pursuing fairness (Murphy et al., 2011; van Lange, 1999; van Lange et al., 1997). Among these, altruistic preference has received substantial research attention. Social value orientation theory defines altruistic preference as "disregarding self-interest and only wishing to maximize others' payoffs" (Murphy et al., 2011). However, in many cases, altruists do not completely ignore self-interest (Pfattheicher et al., 2022; West et al., 2007). Therefore, altruistic preference is more commonly defined as "positive consideration of others' payoffs" (Sáez et al., 2015)—that is, when an individual wishes to increase others' payoffs while holding other factors constant, they are considered to exhibit altruistic preference.

Previous research has extensively explored the causes and mechanisms of altruism, proposing various theoretical explanations. A series of theories suggest that altruism serves a signaling function. For example, the competitive altruism hypothesis (Hardy & Van Vugt, 2006) and costly signaling theory (Gintis et al., 2001) posit that people engage in altruistic behavior competitively to signal their prosocial traits to the group and establish a positive social image, thereby gaining reputation and status that benefit them in subsequent social interactions. In this process, the visibility of altruistic behavior to others is a crucial step for altruism to fulfill its signaling function. This study investigates the influence of behavioral visibility on altruistic behavior from the perspective of altruism' s signaling function.

Some studies have found that in various social interactions, people' s behavior is easily influenced by whether others are aware of or observe their actions (e.g., Dana et al., 2006; Fox & Guyer, 1978; Hamilton & Lind, 2016; Jerdee & Rosen, 1974; Lacetera & Macis, 2010). Bradley et al. (2018) categorized "observability" into three levels of increasing depth: The first is "perceived observation," where visual cues create a sense of being watched without an actual observer present (e.g., Bateson et al., 2006; Burnham & Hare, 2007; Haley & Fessler, 2005; Oda et al., 2011; Raihani & Bshary, 2012). The second is "behavioral observation," where others can know that the behavior occurred but not the actor' s identity (i.e., they do not know personal information such as the actor' s name) (e.g., Andreoni & Bernheim, 2009). The third is "behavioral and identity observation" (e.g., Andreoni & Petrie, 2004; McAuliffe et al., 2013; 2020). Most existing research has focused on the first and third types of observability, while the second type, which has broader ecological validity, remains understudied.

On the other hand, research indicates that the observer' s identity—such as being a recipient (stakeholder) of the altruistic behavior or a third-party observer—may also influence the effect of behavioral visibility (McAuliffe et al., 2020). Most current research focuses on third-party influences on individual altruistic behavior. For instance, studies on "third-party punishment"(Fehr & Fischbacher, 2004; McAuliffe et al., 2015) have found that when allocators know a third party is present, they tend to make more altruistic decisions (Tian, 2016). However, less discussion has addressed how the visibility of behavior to recipients affects altruistic behavior (Andreoni & Bernheim, 2009). Compared to third parties, recipients' interests are directly affected by the allocator' s behavior. Therefore, allocators may exhibit different changes in altruistic behavior when behavior is visible to recipients versus third parties. In the few existing studies on this topic, such as Andreoni and Bernheim (2009), allocators knew whether they determined each trial' s outcome, but recipients only knew the probability that the allocation was determined by the allocator. When allocations were 100% determined by the allocator, they chose options that gave others higher payoffs; when allocations had only a small probability of depending on the allocator' s behavior, their choices became more self-interested. This study demonstrates that recipients' awareness of the allocator' s behavior can increase altruistic behavior. However, it manipulated visibility through probabilities, which is less intuitive and differs substantially from real-world social situations. In summary, distinct from numerous studies examining perceived observation, full identity observation, or third-party observers, this study focuses on how behavioral visibility to recipients influences allocators' altruistic behavior.

To quantify social preferences, this study adopts a social decision-making framework and employs computational modeling to examine the utility functions underlying decisions. The "inequity aversion" model proposed by Fehr and Schmidt (1999) effectively quantifies social preferences and reveals the cognitive mechanisms of social decision-making. Its utility function is expressed as follows:

$$U_i = x_i - \alpha \cdot \max\{x_i - x_j, 0\} - \beta \cdot \max\{x_j - x_i, 0\}$$

This model suggests that individuals consider not only their own net payoff ($U_i = x_i - \alpha \cdot \max\{x_i - x_j, 0\} - \beta \cdot \max\{x_j - x_i, 0\}$) but also the fairness of the allocation (i.e., the difference between self and other payoffs). Specifically, when self payoff exceeds other payoff, individuals exhibit advantageous inequity aversion (AIA, α); when self payoff is less than other payoff, they exhibit disadvantageous inequity aversion (DIA, β). The Fehr-Schmidt model posits that people show fairness preferences in both advantageous and disadvantageous inequity situations (i.e., both AIA and DIA are greater than zero), a finding supported by some studies (e.g., Hu et al., 2018; Sáez et al., 2015). However, other studies have reported cases where DIA and AIA are less than zero (Gao et al., 2018). Sáez et al. (2015) argue that when AIA > 0 and DIA < 0, it indicates that people neither want their payoff to exceed others' nor mind when others' payoff exceeds theirs. In this case, both AIA and DIA can indicate altruistic preference, reflecting a desire to increase others' relative payoffs.

Within the Fehr-Schmidt model framework, combining the positive and negative values of AIA and DIA can represent multiple social preferences across different quadrants (Figure 1 [Figure 1: see original paper]): AIA > 0 and DIA > 0 corresponds to fairness preference; AIA > 0 and DIA < 0 corresponds to altruistic preference (Sáez et al., 2015); AIA < 0 and DIA > 0 indicates acceptance of having more than others while not wanting others to have more, representing a negative relative weight on others' payoffs (non-altruistic preference); AIA $<$ 0 and DIA < 0, in specific experimental designs (Gao et al., 2018), indicates concern for allocation efficiency, with negative AIA representing a desire to increase self payoff and negative DIA representing a desire to increase others' payoff. This study adopts the experimental design of Gao et al. (2018) to define the social preferences corresponding to the positive and negative values of AIA and DIA across four quadrants. Based on this framework, we propose Hypothesis 1: Recipients' visibility of the decision-maker' s behavior can increase the decision-maker' s altruistic preference; compared to the invisible condition, AIA will increase and DIA will decrease in the visible condition.

To quantitatively study social preferences, careful selection of measurement methods is required. Utility theory posits that preferences represent individuals' internal tendencies toward options, which directly influence their decision-making behavior (Raiffa & Schlaifer, 1961). Therefore, by observing individuals' explicit, discrete decision behaviors, we can reverse-engineer and measure their internal, continuous preferences. Regarding social preference measurement, there are two primary behavioral response types: inferring preferences through choice decisions (Gao et al., 2018; Sáez et al., 2015) and measuring preferences through subjective evaluations (ratings) (Liu et al., 2019; Loewenstein et al., 1989). Experiment 1 employs both choice and rating methods to measure social preferences, exploring whether the effect of behavioral visibility on altruistic preference is stable across response types. Accordingly, we propose Hypothesis 2: Behavioral visibility increases altruistic preference in both choice and rating response types.

Previous research on the mechanisms underlying visibility' s influence on altruistic behavior remains limited. Therefore, this study also aims to explore the role of social norms in behavioral visibility' s enhancement of altruistic preference. Social norms are divided into descriptive norms (describing how most people actually behave) and injunctive norms (indicating what most people approve as correct behavior) (Cialdini et al., 1990). Studies have found that conveying descriptive norms to participants can significantly change their behavior, making it converge toward the social norm, both in real-world settings (e.g., Dempsey et al., 2018; Palacios et al., 2022) and laboratory contexts (e.g., Kawamura & Kusumi, 2017; Raihani & McAuliffe, 2014). Thus, this study focuses primarily on descriptive social norms.

Social norms may play an important role in visible social decision-making. First, some researchers argue that social norms largely drive prosocial preferences, finding that different levels of social norm compliance can explain different social preferences (Kimbrough & Vostroknutov, 2016; McBride & Ridinger, 2021). Moreover, one study found that conveying altruistic social norms significantly increased participants' altruistic behavior (Agerström et al., 2016). Second, behavioral visibility and social norms may be connected. For example, when people feel observed, their behavioral variability decreases (Nettle et al., 2013), suggesting that visible behavior may be more likely to conform to specific norms. Some other visibility studies have also mentioned the role of social norms. For instance, Fehr and Fischbacher (2004) argue that third-party presence maintains fairness and cooperation norms. In studies on "perceived observation," providing non-altruistic social norms eliminated the effect of observational cues on donation amounts (Kawamura & Kusumi, 2017). However, in the context of this study—where recipients' visibility of the decision-maker' s behavior increases the decision-maker' s altruistic preference—the influence of social norms remains unknown.

Based on previous research showing that altruistic social norms are widespread in social interactions (Pereda et al., 2017), we hypothesize that behavioral visibility may make participants more aware of such norms, leading to more altruistic behavior. When altruistic norms are absent or diminished, the effect of behavioral visibility on altruistic preference should decrease. Building on this, Experiment 2 investigates and tests Hypothesis 3: The effect of behavioral visibility on increasing altruistic preference is influenced by the presence of altruistic norms. Specifically, the degree to which behavioral visibility increases altruistic preference is greater when altruistic norms exist compared to when they do not.

In summary, this study uses two experiments to test three hypotheses. Both Experiment 1 and Experiment 2 employ a modified dictator game to create social decision-making contexts and use the inequity aversion model to quantify social preferences, investigating the impact of behavioral visibility on social preferences to test Hypothesis 1. Additionally, Experiment 1 measures social preferences through both choice (where participants decide payoffs for self and other) and rating (where participants evaluate given payoff allocations) methods to explore the stability of visibility effects across response types, testing Hypothesis 2. Experiment 2 manipulates social norms to examine how altruistic or non-altruistic norms influence the effect of behavioral visibility on altruistic preference, testing Hypothesis 3.

2 Experiment 1

2.1 Methods

2.1.1 Participants Prior to data collection, we used G*Power 3.1.9.7 to calculate the required sample size. For Experiment 1' s design, based on an F-test for a 2$×$2 repeated-measures ANOVA, with power = 0.8, significance level = 0.05, and effect size = 0.25, the required sample size was 34. In the actual study, we recruited 41 adult university students (16 males). Three participants were excluded due to incomplete data records, leaving 38 valid participants (14 males) aged 19–30 years (M = 22.03, SD = 2.70). All participants were native Chinese speakers, right-handed, and had normal or corrected-to-normal vision. All participants signed informed consent forms before the experiment, were aware of potential benefits and risks, participated voluntarily, and received monetary compensation after the experiment.

2.1.2 Design and Procedure Experiment 1 used a 2 (visibility: visible vs. invisible) × 2 (response type: choice vs. rating) × 2 (inequity type: advantageous vs. disadvantageous) within-subjects design. The independent variables were visibility, response type, and inequity type. The dependent variable was participants' degree of altruism, measured through both choice behavior in the choice condition and subjective ratings in the rating condition.

Before the experiment began, 4-8 same-sex participants arrived at the laboratory simultaneously (in groups of 4, 6, or 8). Participants were told that in the upcoming interpersonal interaction experiment, half of the people present would be randomly designated as allocators and the other half as recipients. Each round of the game would be completed by one allocator randomly paired with one recipient (with re-pairing before each round), and all games would be anonymous. In reality, all participants served as allocators, and the "partners" in the experiment did not actually exist (post-experiment debriefing confirmed that participants believed their partners were real). Participants completed the experiment through a computer program written in MATLAB® R2018a.

The experimental task used a block design with four blocks: visible-choice, visible-rating, invisible-choice, and invisible-rating. The two choice blocks and two rating blocks always appeared together, but the order of choice versus rating blocks was randomized between participants. Within each response type (choice or rating), the order of visible and invisible conditions was also randomized. Before each block, the screen indicated whether the block was visible or invisible and whether participants would make choices or ratings.

In each round, the participant and their "partner" first completed a simple addition/subtraction problem. If both answered correctly, they proceeded to reward allocation and made 4 choices or 4 ratings; otherwise, they waited 20 seconds before moving to the next round. This design aimed to reduce the influence of initial endowment and task contribution on allocation tendencies (Li et al., 2018).

In the choice condition, participants, acting as allocators, selected one of two options to determine the points they and their "partner" would receive in that round. One option was always the fair option (self receives 10 points, partner receives 10 points), while the other was an unfair option (half were advantageous unfair, such as "self receives 12 points, partner receives 8 points," and half were disadvantageous unfair, such as "self receives 8 points, partner receives 12 points"). The option settings were adapted from Gao et al. (2018).

In the rating condition, the screen presented one fair and one unfair allocation scheme, with one selected by the program to determine the participant' s and partner' s points for that round. Participants were told that they and their partner needed to rate their satisfaction with the allocation on a 9-point scale from -4 to 4.

After participants made their responses, they entered the feedback phase. In the visible condition, participants were told that their choice/rating results were being shown to their partner, indicated by an open eye. In the invisible condition, the partner would not see the participant' s choice/rating results, indicated by a closed eye. The trial procedure is illustrated in Figure 2 [Figure 2: see original paper].

The choice blocks consisted of 12 rounds, totaling 12$×4 = 48choices.Theratingblocksconsistedof18rounds, tot
= 72 ratings, of which 48 corresponded to unfair allocation options (matching the choice condition). To make the experimental design appear reasonable, there were also 24 trials where the program selected the fair allocation option.

Finally, all points earned by participants during the experiment were summed and converted to experimental compensation at a 60:1 ratio (each participant received approximately 60 yuan).

2.1.3 Measures The dependent variable was the degree of altruism, measured through both model-free indices of explicit altruistic choices/ratings and computational modeling to reveal implicit altruistic preferences.

For the choice condition, following previous research (Gao et al., 2018; Sáez et al., 2015), we calculated the mean absolute difference between self and other payoffs for the selected options (referred to as "self-other payoff difference"). In advantageous unfair conditions, a smaller self-other payoff difference, and in disadvantageous unfair conditions, a larger self-other payoff difference, indicate higher altruism (less desire to have more than others and greater tolerance for others having more than self). In the rating condition, following previous research (Liu et al., 2019), we used the mean rating for unfair options as the behavioral index: lower ratings for advantageous unfair options and higher ratings for disadvantageous unfair options indicate stronger preferences for advantageous fairness and disadvantageous unfairness, reflecting higher altruism.

For computational modeling, we used the inequity aversion model (Fehr & Schmidt, 1999) to model participants' preferences. We assumed that participants' utility for unfair options in both choice and rating conditions was:

$$U_{\text{unfair}} = M_s - \alpha_1 \cdot \max\{M_s - M_o, 0\} - \beta_1 \cdot \max\{M_o - M_s, 0\}$$

This model assumes that participants' utility consists of self payoff (M_s) and the self-other payoff difference ($M_s - M_o$). Here, α_1 and β_1 represent AIA and DIA in the visible condition, while α_2 and β_2 represent AIA and DIA in the invisible condition. Larger AIA and smaller DIA indicate higher altruism.

For the fair option ($M_s = 10$ and $M_o = 10$), participants' utility was:

$$U_{\text{fair}} = 10$$

In the choice condition, we assumed that participants' selection of the unfair option followed a Bernoulli distribution with parameter P_{unfair}, where P_{unfair} equals the SoftMax transformation of the difference between the two options' subjective utilities:

$$P_{\text{unfair}} = \frac{1}{1 + e^{-\lambda(U_{\text{unfair}} - U_{\text{fair}})}}$$

where λ is the temperature parameter, reflecting participants' sensitivity to decision utility (the difference between unfair and fair option utilities). We assumed different λ values for visible and invisible conditions, denoted as λ_1 and λ_2.

In the rating condition, we assumed a linear relationship between rating scores and the difference between the two options' subjective utilities (Liu et al., 2019). Participants' ratings for unfair and fair options respectively satisfied:

$$R_{\text{unfair}} = b_0 + b_1 \cdot (U_{\text{unfair}} - U_{\text{fair}}) + \epsilon$$

$$R_{\text{fair}} = b_0 + b_1 \cdot (U_{\text{fair}} - U_{\text{unfair}}) + \epsilon$$

where b_0 is the intercept, b_1 is the slope, and ϵ is the error term ($\epsilon \sim N(0, \sigma)$). We assumed different b_0 and b_1 values for visible and invisible conditions, denoted as b_{01}, b_{11} and b_{02}, b_{12}.

We used R 4.0.2 and JAGS 4.3.0 to establish a Hierarchical Bayesian Model (HBM) for parameter estimation. We assumed that individual participants' parameter distributions for visible and invisible conditions overlapped, i.e., $\alpha_2 = \alpha_1 + \Delta\alpha$, $\beta_2 = \beta_1 + \Delta\beta$, $\lambda_2 = \lambda_1 + \Delta\lambda$, $b_{02} = b_{01} + \Delta b_0$, and $b_{12} = b_{11} + \Delta b_1$. In the choice condition, we simultaneously estimated 6 parameters for each participant: $\alpha_1, \beta_1, \lambda_1, \Delta\alpha, \Delta\beta, \Delta\lambda$. In the rating condition, we simultaneously estimated 10 parameters: $\alpha_1, \beta_1, b_{01}, b_{11}, \Delta\alpha, \Delta\beta, \Delta b_0, \Delta b_1, \lambda, \sigma$. In the HBM, individual parameters were drawn from group-level normal distributions. For each parameter, the prior distribution for the mean of the group normal distribution was a normal distribution with parameters (0, 1), and the prior distribution for precision (inverse of variance) was a gamma distribution with parameters (0.1, 0.1). After obtaining the posterior distributions, we used the median of each posterior distribution as the parameter estimate.

We constructed two models for comparison: In the choice-rating joint model, choice and rating conditions shared the same α and β parameter values; in the choice-rating separate model, choice and rating conditions used different α and β parameters.

Before modeling, we verified that in the rating condition, both options influenced participants' ratings. Using a mixed linear model with rating scores as the dependent variable, self and other payoffs in the two options (denoted as $M_{s1}, M_{o1}, M_{s2}, M_{o2}$) and visibility (visible) as fixed effects, and individual random intercepts, we established the following regression model:

$$R_{\text{rating}} = \beta_{0i} + \beta_1 M_{s1} + \beta_2 M_{o1} + \beta_3 M_{s2} + \beta_4 M_{o2} + \beta_5 \text{visible} + u_i + \epsilon$$

where ϵ represents error terms following a standard normal distribution, u_i represents random intercepts, and β coefficients represent the effects of different independent variables.

2.2 Results

We conducted data analysis and statistical testing using IBM SPSS Statistics 20, MATLAB® R2018a, and R 4.0.2, combining model-free results and computational modeling to test our hypotheses.

2.2.1 Model-Free Results We analyzed the effects of behavioral visibility on altruism in choice and rating conditions separately. For the choice condition, we conducted a repeated-measures ANOVA on the absolute self-other payoff difference, with inequity type (advantageous vs. disadvantageous) and visibility (visible vs. invisible) as independent variables. Results showed a significant main effect of inequity type, F(1,37) = 177.59, $p < 0.001$, $\eta_p^2 = 0.83$, with larger self-other payoff differences in advantageous than disadvantageous conditions. The main effect of visibility was not significant, F(1,37) = 2.63, p = 0.113, $\eta_p^2 = 0.066$. The interaction between inequity type and visibility was significant, F(1,37) = 22.00, $p < 0.001$, $\eta_p^2 = 0.37$. Simple main effects analysis revealed that in advantageous conditions, the self-other payoff difference was significantly smaller in the visible than invisible condition (D = -1.07, $p < 0.001$) (Figure 3a [Figure 3: see original paper]); in disadvantageous conditions, the self-other payoff difference was significantly larger in the visible than invisible condition (D = 0.544, p = 0.041) (Figure 3b).

For the rating condition, we conducted a repeated-measures ANOVA on rating scores, with inequity type and visibility as independent variables. Results showed a significant main effect of inequity type, F(1,37) = 84.04, $p < 0.001$, $\eta_p^2 = 0.69$, with higher ratings in advantageous than disadvantageous conditions. The main effect of visibility was not significant, F(1,37) = 1.66, p = 0.205, $\eta_p^2 = 0.043$. The interaction between inequity type and visibility was significant, F(1,37) = 48.81, $p < 0.001$, $\eta_p^2 = 0.57$. Simple main effects analysis revealed that in advantageous conditions, rating scores were significantly lower in the visible than invisible condition (D = -0.43, $p < 0.001$) (Figure 3c); in disadvantageous conditions, rating scores were significantly higher in the visible than invisible condition (D = 0.60, $p < 0.001$) (Figure 3d).

These results demonstrate that visibility' s effect on altruism was consistent across choice and rating conditions: compared to the invisible condition, participants in the visible condition were less tolerant of advantageous inequity and more tolerant of disadvantageous inequity, exhibiting higher overall altruism.

Figure 3 Model-free results from Experiment 1. (a)(b) Choice condition: self-other payoff differences in advantageous and disadvantageous situations. (c)(d) Rating condition: rating scores in advantageous and disadvantageous situations. Horizontal solid lines inside boxplots represent medians, dashed lines represent means. *: p < 0.05;* ***: p < 0.01;*** : p < 0.001.

2.2.2 Computational Modeling Results Before computational modeling, we first verified the reasonableness of our modeling approach. Using equation (6) from the Methods section for regression analysis, we found that participants' ratings were simultaneously influenced by the self and other payoffs of both selected and unselected options ($\beta_{M_{s1}} = 0.18$, t = 55.35, p < 0.001; $\beta_{M_{o1}} = -0.016$, t = -4.51, p < 0.001; $\beta_{M_{s2}} = -0.10$, t = -22.12, p < 0.001; $\beta_{M_{o2}} = -0.030$, t = -6.47, p < 0.001). Therefore, our utility modeling for the rating condition considered both selected and unselected options. Model comparison results showed that the choice-rating separate model (DIC = 21233.43) outperformed the choice-rating joint model (DIC = 22185.17).

Parameter estimates from the choice-rating separate model are shown in Figures 4(a) and (b) [Figure 4: see original paper].

Figure 4 Model-based results from Experiment 1. (a)(b) Model-estimated AIA and DIA. (c)(d) Differences in AIA and DIA between visible and invisible conditions for choice and rating conditions. In (a)(b), significance above each boxplot indicates the parameter is significantly greater than zero, while significance below indicates it is significantly less than zero. Significance at the top of the figure indicates comparison of differences between choice and rating conditions across visibility. Horizontal solid lines inside boxplots represent medians, dashed lines represent means. *: p < 0.05;* ***: p < 0.01;*** : p < 0.001. In (c)(d), light patterns show individual-level model fits, dark patterns show condition means.

To examine differences in AIA and DIA across visibility and response types, we conducted separate 2$×$2 repeated-measures ANOVAs with AIA and DIA as dependent variables and visibility and response type as independent variables.

The ANOVA on AIA revealed a significant main effect of visibility, F(1,37) = 255.17, p < 0.001, $\eta_p^2 = 0.87$, with greater AIA in visible than invisible conditions. The main effect of response type was significant, F(1,37) = 11.15, p = 0.002, $\eta_p^2 = 0.23$, with smaller AIA in choice than rating conditions. The interaction between visibility and response type was significant, F(1,37) = 38.53, p < 0.001, $\eta_p^2 = 0.55$. Simple main effects analysis showed that in invisible conditions, AIA was significantly smaller in choice than rating conditions (D = -0.26, p < 0.001); in visible conditions, AIA was also marginally smaller in choice than rating conditions (D = -0.13, p = 0.032).

The ANOVA on DIA revealed a significant main effect of visibility, F(1,37) = 123.49, p < 0.001, $\eta_p^2 = 0.77$, with smaller DIA in visible than invisible conditions. The main effect of response type was significant, F(1,37) = 22.01, p < 0.001, $\eta_p^2 = 0.37$, with smaller DIA in choice than rating conditions. The interaction between visibility and response type was significant, F(1,37) = 32.13, p < 0.001, $\eta_p^2 = 0.47$. Simple main effects analysis showed that in invisible conditions, DIA was significantly smaller in choice than rating conditions (D = -0.60, p < 0.001); in visible conditions, DIA was also significantly smaller in choice than rating conditions (D = -0.46, p < 0.001).

Figures 4(c-d) visually depict the effects of visibility and response type on AIA and DIA. Compared to invisible conditions, visible conditions showed increased AIA and decreased DIA, shifting overall toward the fourth quadrant and indicating that behavioral visibility enhances altruistic preference (Figure 4c). Compared to rating conditions, choice conditions showed decreased AIA and DIA, shifting overall toward the third quadrant and indicating that participants cared more about allocation efficiency during choice (Figure 4d).

The interaction between visibility and response type in computational modeling results suggests that visibility reduces parameter differences between choice and rating conditions. To further verify this, we calculated the differences in AIA and DIA between choice and rating conditions and conducted paired-sample t-tests on these differences across visible and invisible conditions. Results showed that compared to invisible conditions, visible conditions produced smaller differences between choice and rating conditions for both AIA ($t = -6.21$, $p < 0.001$, Cohen' s $d = 0.37$) and DIA ($t = -5.67$, $p < 0.001$, Cohen' s $d = 0.20$). This indicates that relative to invisible behavior, visible behavior reduces within-individual variability in social preferences between choice and rating conditions, making social preferences more consistent.

3 Experiment 2

Experiment 1 found that visibility reduced preference differences across response types and made preferences more consistent, suggesting that visibility may cause participants to conform more strongly to some social norm. Based on this result, Experiment 2 further manipulated social norms to investigate the role of social norms in the effect of behavioral visibility on altruistic preference.

3.1 Methods

3.1.1 Participants Prior to data collection, we used G*Power 3.1.9.7 to calculate the required sample size. For a 2$×$2 repeated-measures ANOVA with power = 0.8, significance level = 0.05, and effect size = 0.25, the required sample size was 34. In the actual study, we recruited 54 adult university students (23 males). One participant was excluded due to stereotyped responding (always selecting the left option), leaving 53 valid participants (22 males) aged 18–26 years ($M = 20.33$, $SD = 1.82$). All participants were native Chinese speakers, right-handed, and had normal or corrected-to-normal vision. All participants signed informed consent forms before the experiment, were aware of potential benefits and risks, participated voluntarily, and received monetary compensation after the experiment.

3.1.2 Design and Procedure Experiment 2 used a 2 (visibility: visible vs. invisible) × 2 (social norm: altruistic vs. non-altruistic) within-subjects design. The independent variables were visibility and social norm, and the dependent variable was participants' degree of altruism. Experiment 1 confirmed that the effect of behavioral visibility on increasing altruistic preference held in both choice and rating conditions. Therefore, Experiment 2 used only the choice condition.

The visibility manipulation was identical to Experiment 1. For social norms, following the definition of descriptive norms (Elster, 1989), we manipulated altruistic or non-altruistic social norms by presenting "the proportion of previous participants who chose unfair options"(see Figure 5 [Figure 5: see original paper], with gray bars in each option' s background and percentages on the right showing choice proportions). Specifically, we selected 5 points in the fourth quadrant (altruistic preference) and 5 points in the second quadrant (non-altruistic preference) of the coordinate system with AIA on the x-axis and DIA on the y-axis (see Figure 5; altruistic quadrant points: [0.5, -0.5], [0.6, -0.2], [0.2, -0.6], [0.3, -0.4], [0.4, -0.3]; non-altruistic quadrant points: [-0.5, 0.5], [-0.6, 0.2], [-0.2, 0.6], [-0.3, 0.4], [-0.4, 0.3]). Using the choice utility model described in the "Measures" section, we randomly generated 10 sets of virtual participant choice data for each virtual preference combination (i.e., 50 virtual participants per norm condition). We then calculated, for each choice trial, the proportion of these 50 virtual participants who selected the unfair option under each norm. To make the within-subjects manipulation of social norms more plausible, participants were told they would sequentially log into different game servers and see the proportion of previous participants from that server who chose each option. Server 1 corresponded to the altruistic social norm, and Server 2 corresponded to the non-altruistic social norm. Participants were told their choice data would be synchronized to these servers and presented to future participants. This setup placed participants in different social groups and made them aware of the majority behavior patterns in each group.

Experiment 2 also used a block design, with the two independent variables forming 2$\times$2 = 4 blocks: altruistic-visible, altruistic-invisible, non-altruistic-visible, and non-altruistic-invisible. Blocks with the same social norm appeared consecutively, with the order of different social norms randomized between participants. Within each social norm, the order of visible and invisible blocks was randomized. Before each block, the screen indicated the visibility condition and server type.

The procedure for Experiment 2 was similar to the choice task in Experiment 1, with the addition of social norm manipulation. To ensure participants formed a stable impression of the current social norm, each block began with 3 practice rounds, followed by 12 formal rounds (48 binary choices total). Each round consisted of an addition/subtraction problem and 4 consecutive monetary allocation tasks. Specific option settings for Experiment 2 are available in our online open data repository (https://github.com/psych575/open-data-and-code-for-xb21-575.git).

Figure 5 Experiment 2 procedure

3.1.3 Measures Similar to Experiment 1, Experiment 2 measured altruism through both model-free indices and model-based parameters.

For model-free indices, to more intuitively represent altruism, we used the average payoff participants allocated to others across the four conditions, termed "average points given to others." Higher values indicate greater altruism. Additionally, following Experiment 1' s choice condition, we calculated the mean absolute self-other payoff difference for selected options in advantageous and disadvantageous situations across the four conditions.

We modeled participants' behavior using the inequity aversion model (Fehr & Schmidt, 1999). We assumed participants' utility for unfair options was:

$$U_{\text{unfair}} = M_s - \alpha \cdot \max\{M_s - M_o, 0\} - \beta \cdot \max\{M_o - M_s, 0\}$$

We estimated AIA and DIA separately for the four conditions (altruistic-visible, altruistic-invisible, non-altruistic-visible, non-altruistic-invisible), with corresponding AIA values $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and DIA values $\beta_1, \beta_2, \beta_3, \beta_4$. Other utility model settings were identical to Experiment 1' s choice condition. This model was termed the altruistic-non-altruistic separate model. Additionally, we assumed parameters were identical across social norms, with AIA and DIA affected only by visibility, creating an altruistic-non-altruistic joint model.

We used R 4.0.2 and JAGS 4.3.0 to establish a Hierarchical Bayesian Model (HBM) for parameter estimation. We assumed that AIA, DIA, and temperature parameters across conditions overlapped, i.e., $\alpha_2 = \alpha_1 + \Delta\alpha_2$, $\alpha_3 = \alpha_1 + \Delta\alpha_3$, $\alpha_4 = \alpha_1 + \Delta\alpha_4$, $\beta_2 = \beta_1 + \Delta\beta_2$, $\beta_3 = \beta_1 + \Delta\beta_3$, $\beta_4 = \beta_1 + \Delta\beta_4$, $\lambda_2 = \lambda_1 + \Delta\lambda_2$, $\lambda_3 = \lambda_1 + \Delta\lambda_3$, $\lambda_4 = \lambda_1 + \Delta\lambda_4$. For each participant, we simultaneously estimated 12 parameters: $\alpha_1, \beta_1, \lambda_1, \Delta\alpha_2, \Delta\alpha_3, \Delta\alpha_4, \Delta\beta_2, \Delta\beta_3, \Delta\beta_4, \Delta\lambda_2, \Delta\lambda_3, \Delta\lambda_4$. Other HBM settings were identical to Experiment 1' s choice condition.

3.2 Results

We conducted data analysis and statistical testing using IBM SPSS Statistics 20, MATLAB® R2018a, and R 4.0.2, combining model-free results and computational modeling to test our hypotheses.

3.2.1 Model-Free Results First, we conducted a 2$×$2 repeated-measures ANOVA on average points given to others, with visibility and social norm as independent variables. Results showed a significant main effect of visibility, F(1,52) = 12.38, p = 0.001, $\eta_p^2 = 0.19$, with more points given to others in visible than invisible conditions. The main effect of social norm was significant, F(1,52) = 53.74, p < 0.001, $\eta_p^2 = 0.53$, with more points given to others under altruistic than non-altruistic norms. The interaction between visibility and social norm was marginally significant, F(1,52) = 3.09, p = 0.085, $\eta_p^2 = 0.06$. Simple main effects analysis revealed that under altruistic norms, participants gave significantly more points to others in visible than invisible conditions (D = 0.37, p = 0.001); under non-altruistic norms, there was no significant difference in points given between visible and invisible conditions (D = 0.12, p = 0.185) (Figure 6c [Figure 6: see original paper]). This directly demonstrates that visibility' s effect on altruism depends on altruistic social norms.

Second, we conducted a 2$×$2 repeated-measures ANOVA on self-other payoff differences in advantageous conditions, with visibility and social norm as independent variables. Results showed a significant main effect of visibility, F(1,52) = 32.96, $p < 0.001$, $\eta_p^2 = 0.39$, with smaller self-other payoff differences in visible than invisible conditions. The main effect of social norm was significant, F(1,52) = 16.95, $p < 0.001$, $\eta_p^2 = 0.25$, with smaller self-other payoff differences under altruistic than non-altruistic norms. The interaction between visibility and social norm was significant, F(1,52) = 5.04, p = 0.029, $\eta_p^2 = 0.09$. Simple main effects analysis revealed that under altruistic norms, self-other payoff differences were significantly smaller in visible than invisible conditions (D = -0.68, $p < 0.001$), indicating less tolerance for advantageous inequity when visible; under non-altruistic norms, visible conditions also showed significantly smaller self-other payoff differences (D = -0.34, p = 0.001) (Figure 6a).

We conducted an ANOVA on absolute self-other payoff differences in disadvantageous conditions with visibility and social norm as independent variables. Results showed a marginally significant main effect of visibility, F(1,52) = 3.08, p = 0.085, $\eta_p^2 = 0.06$, with larger self-other payoff differences in visible than invisible conditions. The main effect of social norm was significant, F(1,52) = 42.38, $p < 0.001$, $\eta_p^2 = 0.45$, with larger self-other payoff differences under altruistic than non-altruistic norms. The interaction between visibility and social norm was not significant, F(1,52) = 2.01, p = 0.162, $\eta_p^2 = 0.04$. However, paired-sample t-tests revealed that under altruistic norms, self-other payoff differences were marginally significantly larger in visible than invisible conditions (t = 2.06, p = 0.088, Bonferroni-corrected, Cohen's d = 0.23), indicating greater tolerance for disadvantageous inequity when visible; under non-altruistic norms, there was no significant difference in self-other payoff differences between visible and invisible conditions (t = 0.44, p = 0.663, Cohen' s d = 0.04) (Figure 6b).

Figure 6 Results from Experiment 2. (a)(b) Self-other payoff differences in advantageous and disadvantageous situations. (c) Average points given to others. (d)(e) AIA and DIA estimated from hierarchical Bayesian models. Horizontal solid lines inside boxplots represent medians, dashed lines represent means. In (d)(e), significance above each boxplot indicates the parameter is significantly greater than zero, while significance below indicates it is significantly less than zero. *: $p < 0.05$;* **: $p < 0.01$;** : $p < 0.001$.

3.2.2 Computational Modeling Results Model comparison showed that the altruistic-non-altruistic separate model (DIC = 4374.87) outperformed the altruistic-non-altruistic joint model (DIC = 4680.11). Therefore, we used the altruistic-non-altruistic separate model for subsequent analysis. Parameter estimates from this model are shown in Figures 6(d) and (e).

To examine the effects of visibility and social norm on AIA and DIA, we conducted separate 2$×$2 repeated-measures ANOVAs with AIA and DIA as dependent variables and visibility and social norm as independent variables. Our primary focus was the interaction between visibility and social norm, to investigate whether visibility' s effects on AIA and DIA were moderated by social norms.

The ANOVA on AIA revealed a significant main effect of visibility, F(1,52) = 86.87, p $<$ 0.001, $\eta_p^2 = 0.63$, with greater AIA in visible than invisible conditions. The main effect of social norm was significant, F(1,52) = 39.47, p $<$ 0.001, $\eta_p^2 = 0.43$, with greater AIA under altruistic than non-altruistic norms. The interaction between visibility and social norm was marginally significant, F(1,52) = 3.82, p = 0.056, $\eta_p^2 = 0.068$. Simple main effects analysis revealed that under altruistic norms, AIA was greater in visible than invisible conditions (D = 0.10, p $<$ 0.001); under non-altruistic norms, AIA was also greater in visible than invisible conditions (D = 0.063, p $<$ 0.001).

The ANOVA on DIA revealed a significant main effect of visibility, F(1,52) = 4.82, p = 0.033, $\eta_p^2 = 0.085$, with smaller DIA in visible than invisible conditions. The main effect of social norm was significant, F(1,52) = 59.96, p $<$ 0.001, $\eta_p^2 = 0.54$, with smaller DIA under altruistic than non-altruistic norms. The interaction between visibility and social norm was not significant, F(1,52) = 2.55, p = 0.116, $\eta_p^2 = 0.047$. However, paired-sample t-tests revealed that under altruistic norms, DIA was significantly smaller in visible than invisible conditions (t = -2.97, p = 0.008, Bonferroni-corrected, Cohen' s d = 0.26), indicating greater tolerance for disadvantageous inequity when visible; under non-altruistic norms, there was no significant difference in DIA between visible and invisible conditions (t = -0.51, p = 0.612, Cohen' s d = 0.034).

These computational modeling results correspond with model-free results: under altruistic norms, visible conditions showed increased AIA and decreased DIA compared to invisible conditions; under non-altruistic norms, visibility' s effects on AIA and DIA were diminished, indicating that behavioral visibility' s enhancement of altruistic preference depends on altruistic social norms.

4 Discussion

This study used behavioral experiments to construct social decision-making scenarios, investigating how recipients' visibility of allocators' behavior influences allocators' altruistic preference in a modified two-person dictator game. Combining computational modeling with the inequity aversion model to quantify social preferences, we examined changes in advantageous and disadvantageous inequity aversion to reveal shifts in altruistic preference across contexts. Experiment 1 used both choice and rating methods to measure social preferences, finding that regardless of response type, participants showed higher altruism when their behavior was visible to recipients than when invisible, demonstrating that behavioral visibility increases altruistic preference with consistent effects across response types. Additionally, participants showed greater concern for allocation efficiency in the choice condition than in the rating condition. Experiment 2 manipulated social norms to examine how specific norms influence the effect of behavioral visibility on altruistic preference, finding an interaction between social norms and visibility: the degree to which visibility increased altruistic preference was greater under altruistic norms than under non-altruistic norms.

Our findings provide experimental evidence supporting the signaling function hypothesis of altruistic behavior. Based on competitive altruism hypothesis and costly signaling theory, which posit that people are altruistic to signal their prosocial traits and build positive social images (Gintis et al., 2001; Hardy & Van Vugt, 2006), Hardy and Van Vugt (2006) only verified the effect of behavioral visibility on altruistic behavior in public goods games. Our study extends the visibility manipulation to two-person dictator games, finding that when allocation behavior is visible to recipients, allocators choose fewer advantageous options and more disadvantageous options, exhibiting more altruistic behavior overall. Conversely, when behavior is invisible, altruistic behavior decreases and even shows self-interested tendencies: in Experiment 1' s choice condition, AIA values were negative when invisible, indicating a preference for advantageous inequity. These results support strategic motivations for altruism (Böckler et al., 2016) and suggest that people strategically reduce altruistic behavior and increase self-payoff when behavior is invisible. Our study integrates theories from social psychology explaining altruism with social preference research from social decision-making, providing insights into the functions and motivations of altruism.

Furthermore, our study fills gaps in observability research. Many studies have examined "perceived observation," "identity observation," and "third-party observation" (Bradley et al., 2018), but few have investigated "behavioral observation" in decision contexts where behavior is visible to recipients. A few studies have explored similar topics, but their visibility manipulations were less intuitive (Andreoni & Bernheim, 2009). Our study uses interpersonal interaction contexts to manipulate whether behavior is visible to recipients, and both Experiment 1 and Experiment 2 demonstrate that behavioral visibility promotes altruistic preference, providing robust evidence for this conclusion.

It is important to note that our study did not quantitatively manipulate or examine specific differences between "behavior visible to recipients" and other observability types. Regarding the depth of observation, Bradley et al. (2018) found no significant difference in the overall effect size of observability on prosocial behavior across "perceived observation," "behavioral observation," and "behavioral and identity observation." However, different observability types may involve different psychological mechanisms. For example, perceived observation using eye images is influenced by arousal level and stimulus duration (Hesslinger et al., 2017), while behavioral observation involves processing the fact that "my behavior is seen by others," likely engaging more social cognitive processes. Identity-visible interactions occur more among acquaintances and may involve more reciprocity. Therefore, it remains necessary to distinguish different levels of observation in research and interpret mechanisms cautiously. Regarding observer identity, Bradley et al. (2018) compared "disinterested third-party observation" and "other experimental participants observation," finding that third-party observation promoted allocators' altruistic behavior more than other participants' observation. However, experimental participants could serve as recipients, involving greater stake in the allocation, more concern for outcomes, and greater potential for reciprocity, which should produce stronger altruism-promoting effects than third-party observation. Such differences may arise because the "other experimental participants observation" included in Bradley et al. (2018) were not strictly controlled studies of "recipient visibility." Our study fills this gap, providing more empirical evidence for future research exploring differences across observer identities.

Additionally, Experiment 1 found that response type influenced social preferences. Compared to rating conditions, choice conditions showed decreased AIA and DIA, indicating greater concern for allocation efficiency. Choshen-Hillel and Yaniv (2011; 2012) argue that when making choices, individuals have decision power and can actually influence payoffs for self and others, leading to greater concern for allocation efficiency (i.e., maximizing total payoff). In rating tasks, individuals focus more on fairness and are more satisfied with smaller self-other payoff differences. Our results are largely consistent with this view.

Experiment 1 also found an interaction between behavioral visibility and response type: visibility reduced social preference differences between choice and rating conditions, suggesting that visibility makes individuals' preferences and strategies more consistent across contexts. This result aligns with Nettle et al. (2013), who found that feeling observed reduces variability in social behavior. We infer that this reduction in behavioral variability likely reflects visibility's role in making people follow norms more closely. In human society, altruism is a widely accepted social norm (Pereda et al., 2017). Therefore, when behavior is visible to others, people may follow altruistic social norms more closely, leading to more altruistic behavior.

To test this hypothesis, Experiment 2 manipulated social norms by imposing altruistic or non-altruistic norms. Results showed that when altruistic norms were present, behavioral visibility significantly increased altruistic preference; when non-altruistic norms were present, this effect diminished or disappeared. This demonstrates that behavioral visibility's enhancement of altruistic preference depends on the presence of altruistic social norms, providing insight into the mechanism underlying this effect. Additionally, our study found a substantial main effect of social norms, confirming the strong influence of social norms on social preferences and reaffirming the close relationship between social norms and social preferences (Kimbrough & Vostroknutov, 2016).

By contrast, some studies on "perceived observation" have attempted to explore whether social norms explain the effect of feeling observed on altruistic behavior

(Fathi et al., 2014; Kawamura & Kusumi, 2017; Oda & Ichihashi, 2016), but have not reached definitive conclusions. Most of these studies used donation tasks with eye images on donation boxes to create a sense of being observed, manipulating social norms by changing the amount of money already in the donation box. Results showed that when the donation box contained larger denominations (altruistic norm), eye images did not significantly increase donation amounts more than when it contained smaller denominations (non-altruistic norm) (Fathi et al., 2014; Oda & Ichihashi, 2016). Other studies found effects of social norms that were not replicable (Kawakami & Kusumi, 2017). These findings suggest that the effect of perceived observation on altruistic behavior is difficult to explain through social norms. In contrast, our study found that social norms significantly moderate the effect of behavioral visibility on altruistic preference, suggesting that when behavior is visible, participants conform more to altruistic social norms and thus behave more altruistically. This difference may stem from psychological process differences between "behavior visible to recipients" and "perceived observation" : perceived observation is related to arousal levels (Hesslinger et al., 2017) and may be less influenced by social factors like social norms, whereas behavioral visibility may involve more cognitive processing of the social situation, better integrating social norm information into the decision process.

Our study also has limitations and room for improvement. By manipulating social norms, we found that social norms moderate the effect of behavioral visibility on altruistic preference, but we do not exclude other possible mechanisms such as reputation and empathy. Some studies have highlighted the important role of reputation in promoting altruistic and prosocial behavior (Hardy & Van Vugt, 2006; Roberts, 1998; Suzuki & Akiyama, 2005; Sylwester & Roberts, 2010; Van Vugt et al., 2012; Wedekind & Braithwaite, 2002). When behavior is visible, people accumulate reputation through altruistic behavior to build positive social images (Milinski et al., 2002). In contrast, when behavior is invisible, altruistic behavior cannot enhance reputation. Therefore, behavioral visibility may promote altruistic behavior through reputation-seeking. While reputation and social norm compliance may differ in experimental manipulation and interpretation (e.g., Piazza & Bering, 2008), both reflect social desirability—people' s tendency to behave in socially expected ways. This also reflects altruism' s signaling function: decision-makers signal their prosociality through altruistic behavior. Furthermore, the effect of recipient visibility in our study may also relate to empathy and theory of mind (Brüne & Brüne-Cohrs, 2006; Elliott et al., 2011). When recipients can see the decision-maker' s behavior, decision-makers may become more aware of recipients' situations, experience their emotions, and take their perspective, leading to more altruistic behavior. Previous research has shown that both empathy and theory of mind relate to social preferences (e.g., Tsoi & McAuliffe, 2020; Wu & Han, 2021), but no studies have examined their roles in behavioral visibility' s enhancement of altruistic preference—an avenue for future research.

Additionally, our study used choice and rating methods to measure social preferences, assuming both behaviors are based on the same underlying utility function form to computationally model social preferences and compare model parameters across response types. Our comparison builds on widely used social utility function models, with interpretations based on experimental design and estimated parameters. Future research could explore other suitable quantitative indices and comparison methods.

References

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5), 1607–1636.

Andreoni, J., & Petrie, R. (2004). Public goods experiments without confidentiality: A glimpse into fund-raising. *Journal of Public Economics*, 88(7), 1605–1623.

Agerström, J., Carlsson, R., Nicklasson, L., & Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, 52, 147–153.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2(3), 412–414.

Böckler, A., Tusche, A., & Singer, T. (2016). The structure of human prosociality: Differentiating altruistically motivated, norm motivated, strategically motivated, and self-reported prosocial behavior. *Social Psychological and Personality Science*, 7(6), 530–541.

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences*, 285(1875), 20180116.

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience and Biobehavioral Reviews*, 30(4), 437–455.

Burnham, T. C., & Hare, B. (2007). Engineering human cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature*, 18(2), 88–108.

Choshen-Hillel, S., & Yaniv, I. (2011). Agency and the construction of social preference: Between inequality aversion and prosocial behavior. *Journal of Personality and Social Psychology*, 101(6), 1253–1261.

Choshen-Hillel, S., & Yaniv, I. (2012). Social preferences shaped by conflicting motives: When enhancing social welfare creates unfavorable comparisons for the self. *Judgment and Decision Making*, 7(5), 618–627.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don' t know won' t hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2), 193–201.

Dempsey, R. C., McAlaney, J., & Bewick, B. M. (2018). A critical appraisal of the social norms approach as an interventional strategy for health-related behavior and attitude change. *Frontiers in Psychology*, 9, 2180.

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, 48(1), 43–49.

Elster, J. (1989). Social norms and economic theory. *The Journal of Economic Perspectives*, 3(4), 99–117.

Fathi, M., Bateson, M., & Nettle, D. (2014). Effects of watching eyes and norm cues on charitable giving in a surreptitious behavioral experiment. *Evolutionary Psychology*, 12(5), 878–887.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63–87.

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

Fox, J., & Guyer, M. (1978). "Public" choice and cooperation in n-person prisoner' s dilemma. *The Journal of Conflict Resolution*, 22(3), 469–481.

Gao, X., Yu, H., Sáez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion. *Proceedings of the National Academy of Sciences*, 115(33), E7680–E7689.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology*, 213(1), 103–119.

Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision making and the brain* (2nd ed.). Academic Press.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody' s watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26(3), 245–256.

Hamilton, A., & Lind, F. (2016). Audience effects: What can they tell us about social neuroscience, theory of mind and autism? *Culture and Brain*, 4(2), 159–177.

Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, 32(10), 1402–1413.

Hesslinger, V. M., Carbon, C., & Hecht, H. (2017). The sense of being watched is modulated by arousal and duration of the perceptual episode. *i-Perception*, 8(6), 2041669517742179.

Hu, Y., He, L., Zhang, L., Wölk, T., Dreher, J., & Weber, B. (2018). Spreading inequality: Neural computations underlying paying-it-forward reciprocity. *Social Cognitive and Affective Neuroscience*, 13(6), 578-589.

Jerdee, T. H., & Rosen, B. (1974). Effects of opportunity to communicate and visibility of individual decisions on behavior in the common interest. *Journal of Applied Psychology*, 59(6), 712-716.

Kawamura, Y., & Kusumi, T. (2017). The norm-dependent effect of watching eyes on donation. *Evolution and Human Behavior*, 38(5), 659-666.

Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608-638.

Lacetera, N., & Macis, M. (2010). Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior & Organization*, 76(2), 225-237.

Li, O., Xu, F., & Wang, L. (2018). Advantageous inequity aversion does not always exist: The role of determining allocations modulates preferences for advantageous inequity. *Frontiers in Psychology*, 9, 749-749.

Liu, Y., Li, S., Lin, W., Li, W., Yan, X., Wang, X., Pan, X., Rutledge, R. B., & Ma, Y. (2019). Oxytocin modulates social value representations in the amygdala. *Nature Neuroscience*, 22(4), 633-641.

Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 57(3), 426-441.

McAuliffe, K., Blake, P. R., Kim, G., Wrangham, R. W., & Warneken, F. (2013). Social influences on inequity aversion in children. *PloS One*, 8(12), e80966.

McAuliffe, K., Blake, P. R., & Warneken, F. (2020). Costly fairness in children is influenced by who is watching. *Developmental Psychology*, 56(4), 773-782.

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, 134, 1-10.

McBride, M., & Ridinger, G. (2021). Beliefs also make social-norm preferences social. *Journal of Economic Behavior & Organization*, 191(3), 765-784.

Milinski, M., Semmann, D., & Krambeck, H. (2002). Reputation helps solve the ‘tragedy of the commons’. *Nature*, 415(6870), 424-426.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771-781.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., & Bateson, M. (2013). The watching eyes effect in the dictator game: It' s not how much you give, it' s being seen to give something. *Evolution and Human Behavior*, 34(1), 35–40.

Oda, R., & Ichihashi, R. (2016). Effects of eye images and norm cues on charitable donation: A field experiment in an izakaya. *Evolutionary Psychology*, 14(4), 147470491666887.

Oda, R., Niwa, Y., Honma, A., & Hiraishi, K. (2011). An eye-like painting enhances the expectation of a good reputation. *Evolution and Human Behavior*, 32(3), 166–171.

Palacios, J., Fan, Y., Yoeli, E., Wang, J., Chai, Y., Sun, W., Rand, D. G., & Zheng, S. (2022). Encouraging the resumption of economic activity after COVID-19: Evidence from a large-scale field experiment in China. *Proceedings of the National Academy of Sciences*, 119(5), e2100719119.

Pereda, M., Brañas-Garza, P., Rodríguez-Lara, I., & Sánchez, A. (2017). The emergence of altruism as a social norm. *Scientific Reports*, 7(1), 9684.

Pfattheicher, S., Nielsen, Y. A., & Thielmann, I. (2022). Prosocial behavior and altruism: A review of concepts and definitions. *Current Opinion in Psychology*, 44, 124–129.

Piazza, J., & Bering, J. M. (2008). Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behavior*, 29(3), 172–178.

Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. University College London: Biometrika Office.

Raihani, N. J., & Bshary, R. (2012). A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proceedings of the Royal Society B: Biological Sciences*, 279(1742), 3556–3564.

Raihani, N. J., & McAuliffe, K. (2014). Dictator game giving: The importance of descriptive versus injunctive norms. *PloS One*, 9(12), e113826.

Roberts, G. (1998). Competitive altruism: From reciprocity to the handicap principle. *Proceedings of the Royal Society B: Biological Sciences*, 265(1394), 427–431.

Sáez, I., Zhu, L., Set, E., Kayser, A., & Hsu, M. (2015). Dopamine modulates egalitarian behavior in humans. *Current Biology*, 25(7), 912–919.

Suzuki, S., & Akiyama, E. (2005). Reputation and the evolution of cooperation in sizable groups. *Proceedings of the Royal Society B: Biological Sciences*, 272(1570), 1373–1377.

Sylwester, K., & Roberts, G. (2010). Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters*, 6(5), 659–662.

Tian, Y. (2016). *Research on game experiments of fairness preference: The influence of dominance status, punishment situation and fairness perception on fairness preference* [Master' s thesis, Harbin Engineering University]. Harbin.

Tsoi, L., & McAuliffe, K. (2020). Individual differences in theory of mind predict inequity aversion in children. *Personality and Social Psychology Bulletin*, 46(4), 559–571.

van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77(2), 337–349.

van Lange, P. A. M., de Bruin, E. M. N., Otten, W., & Joireman, J. A. (1997). The development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, 73(4), 733–746.

Van Vugt, M., Roberts, G., & Hardy, C. (2012). Competitive altruism: A theory of reputation-based cooperation in groups. In L. Barrett & R. Dunbar (Eds.), *Oxford Handbook of Evolutionary Psychology* (pp. 531–540). Oxford University Press.

Wedekind, C., & Braithwaite, V. A. (2002). The long-term benefits of human generosity in indirect reciprocity. *Current Biology*, 12(12), 1012–1015.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20(2), 415–432.

Wu, T., & Han, S. (2021). Neural mechanisms of modulations of empathy and altruism by beliefs of others' pain. *eLife*, 10, e66043.

Note: Figure translations are in progress. See original paper for figures.