Prediction Bias in Altruistic Display: Helpers Underestimate Observers' Social Evaluation

Authors: Wang Tianhong, Jin Shan, Cheng Zipeng, Lou Yu, Xie Xiaofei, Xie Xiaofei

Date: 2024-01-31T00:00:00+00:00

Abstract

Helpers often face a dilemma when deciding whether to proactively display their altruistic behavior, as they both hope to gain appreciation from others and fear incurring negative evaluation. Does showing off altruistic behavior truly elicit negative evaluation? Are helpers' predictions of social evaluation accurate? The present study, through 7 experiments (N = 1362), reveals the prediction bias in altruistic showing off and its underlying mechanism: for altruistic showing off behavior, helpers underestimate social evaluation from observers; perceived altruistic motivation plays a mediating role in this process. This finding helps deepen the understanding of altruistic showing off behavior and its social evaluation, and provides a basis for encouraging helpers to proactively disseminate altruistic behavior and thereby enhance the altruistic atmosphere at the societal level.

Full Text

The Prediction Bias in Conspicuous Altruism: Helpers Underestimate Social Evaluations from Bystanders

WANG Tianhong, JIN Shan, CHENG Zipeng, LOU Yu, XIE Xiaofei
School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China

Abstract

Helpers often face a dilemma when deciding whether to actively display their altruistic actions, desiring recognition yet fearing negative evaluations. Does showcasing altruistic behavior truly invite criticism? Are helpers accurate in predicting social evaluations? Through seven experiments (N = 1,362), this study reveals the prediction bias in conspicuous altruism and its underlying

mechanisms: helpers underestimate bystanders' social evaluations of conspicuous altruistic behavior, with perceived altruistic motives mediating this process. These findings deepen our understanding of conspicuous altruism and its social evaluations, providing a basis for encouraging helpers to actively disseminate altruistic actions and thereby enhance the culture of altruism at the societal level.

Keywords: conspicuous altruism, prediction bias, impression management, prosocial behavior

1 Introduction

If you participated in voluntary blood donation, would you post your donation certificate on social media? If you saw someone else sharing such a certificate, how would you evaluate them? Altruistic behavior brightens human society, and promoting accessible acts of kindness helps foster a culture of mutual assistance. Simultaneously, helpers can enhance their social image and gain external recognition by displaying their altruistic actions.

However, the moral norm of "doing good without seeking recognition" seems to imply that silent virtue is the only genuine altruism. Such social norms may place helpers in a dilemma, torn between wanting praise and fearing criticism. This "conspicuous altruism paradox"manifests at two levels: individually, helpers wish to publicly display their good deeds yet worry about negative evaluations; societally, we should encourage the active dissemination of altruistic acts, yet negative perceptions of self-promotion may objectively inhibit the spread of altruism.

Does flaunting altruistic behavior truly elicit negative evaluations from others? Addressing this dilemma, the present research focuses on the perspective gap between helpers and bystanders, examining bystanders' actual evaluations of conspicuous altruists and helpers' prediction biases regarding these evaluations, thereby seeking solutions to the conspicuous altruism paradox.

1.1 Conspicuous Altruism and Social Evaluation

This research centers on conspicuous altruism—the intention and behavior of actively displaying one' s good deeds. West (2004) introduced the concept of "conspicuous" from consumer and economic behavior (Veblen & Phillips, 1899) into the prosocial domain, proposing "conspicuous compassion," where individuals deliberately and overtly express sympathy to gain reputation, not from genuine kindness but to appear kind. Subsequent researchers developed related concepts including conspicuous donation (Grace & Griffin, 2006), conspicuous benevolence (Griskevicius et al., 2007), conspicuous generosity (Ellingsen & Johannesson, 2011), conspicuous prosocial consumption (C. M. Johnson et al., 2018), and conspicuous environmentalism (Griskevicius et al., 2010). Domestic scholars have also examined conspicuous prosocial behavior, focusing on its signaling attributes and conspicuous characteristics (姚琦 et al., 2020). Building

on this literature, we define conspicuous altruism as the active, publicly visible display of one' s completed altruistic acts. Since altruistic behavior must occur before it can be displayed, our definition excludes cases where individuals claim or feign altruism without having acted.

Altruistic behavior can transmit positive signals and yield potential benefits. According to costly signaling theory, behavioral costs reliably signal individual traits (Zahavi, 1975). Altruism is an observable behavior that conveys altruistic traits and abilities (Miller, 2007; Zahavi & Zahavi, 1997), often requiring costly investment, thus meeting three key criteria for effective signals: observability, high trait correlation, and difficulty to fake (S. G. B. Johnson & Park, 2021).

People may recognize altruism' s signaling value and employ it as an impression management strategy (Green & Peloza, 2014; Karlan & McConnell, 2014). Research shows that public contexts increase helping frequency (Lacetera & Macis, 2010) and intensity (Van Vugt & Hardy, 2010), while mating motives enhance women' s preference for public benevolence (Griskevicius et al., 2007) and men' s public charitable donations (Iredale et al., 2008). In public settings, individuals respond more positively to other-oriented appeals, with impression management motives mediating this effect (Green & Peloza, 2014).

Anticipated social evaluations directly influence helpers' decisions, representing a crucial factor in choosing how to help and whether to display it. Opportunities for good reputation effectively promote altruism in the real world (Barclay, 2012). Helpers often want their good deeds witnessed, yet fear that self-promotion reveals self-interested motives like image management (Binabou & Tirole, 2021). Altruistic behavior carries unique moral attributes that directly affect social evaluation and reputation (Basu, 2021; S. G. B. Johnson & Park, 2021). When helpers foresee potential negative inferences from conspicuous altruism, they experience conflict and may reduce altruistic behavior (Ariely et al., 2009; Blakely et al., 2003; Danheiser & Graziano, 1982; Gneezy et al., 2010; Kraut, 1973; Newman & Shen, 2012; White & Peloza, 2009). Therefore, understanding both bystanders' actual evaluations and helpers' predictions of these evaluations is essential.

1.2 Perspective Anchoring and Prediction Bias in Conspicuous Altruism

Prediction bias frequently occurs in helping contexts (王逸璐, 谢晓非, 2019) and social interactions generally. As social beings, our actions undergo constant evaluation, affecting interpersonal outcomes and shaping our behavior, including impression management (e.g., Bourdage et al., 2018). However, existing research primarily examines prediction bias between help-seekers and helpers (e.g., Bohns & Flynn, 2010; Newark et al., 2017), with limited investigation into helpers' accuracy in predicting bystanders' evaluations.

In conspicuous altruism, judgments about underlying motives significantly influence social evaluation. The drivers of conspicuous altruism may include purely

selfless altruistic motives (Batson, 2010; Haidt, 2003) aimed at promoting and encouraging altruism (Ashraf et al., 2023; Karlan & McConnell, 2014), or self-oriented factors (Baumeister, 1982; Small & Cryder, 2016), particularly impression management or flaunting motives to gain recognition, respect, and better social image (Lacetera & Macis, 2010; Newman & Cain, 2014), as well as internal benefits like enhanced self-worth (Carlo & Randall, 2002) and positive emotional experiences (Wallace et al., 2020).

Perceived altruistic and flaunting motives substantially impact social evaluation (Berman et al., 2015). Although both "pure"altruistic and "impure"self-interested motives can promote altruistic behavior, social evaluations of these motives differ dramatically. Altruistic and self-interested motives may be perceived as "incompatible" (Lin-Healy & Small, 2012; Newman & Shen, 2012). Consequently, people want others to see their altruistic acts and motives but not infer their flaunting (self-interested) motives. We propose that in conspicuous altruism, both perceived altruistic and flaunting motives may mediate the effect of perspective differences on social evaluations and their predictions. Because helpers and bystanders have different perspectives, information, and reference points, their perceived intensities of flaunting (self-interested) and altruistic motives may differ, leading to discrepancies between bystanders' actual evaluations and helpers' predictions.

Specifically, for helpers, the altruistic act is completed and known; the current concern is the consequences of displaying it. Deliberate imagination increases availability (李燕 et al., 2015), making perceived flaunting motives more salient. Helpers compare themselves to a version who helped without displaying it. Research suggests that when comparing "helpers who display after helping" with "helpers who do not display," people evaluate the latter more positively (Berman et al., 2015; Heyman et al., 2014). Helpers also worry that displaying will reduce others' perception of their altruistic motives (Newman & Cain, 2014). In public contexts, monetary self-interest appeals backfire (Ariely et al., 2009), and people avoid displaying good deeds to prevent revealing flaunting motives (Yang & Hsee, 2022). Thus, helpers predict that bystanders will perceive strong flaunting motives, weak altruistic motives, and consequently give low evaluations.

For bystanders, the helper' s altruistic act is unknown and must be inferred from the displayed content. Without prior knowledge of the helping act, bystanders evaluate self-promoting helpers more positively, perceiving stronger altruistic motives (Berman et al., 2015). When evaluating conspicuous altruism, bystanders compare it to group averages or other self-promotional behaviors, making the altruistic content more salient. This leads to higher perceived altruistic motives, lower perceived self-interest, and more positive evaluations.

Due to this perspective difference, when helpers make predictions, they anchor on their own perspective and adjust insufficiently. Since others' thoughts are not directly accessible, people anchor on their own perspective when inferring others' views, but this adjustment is typically inadequate (Epley et al., 2004; Tamir & Mitchell, 2013), causing prediction bias. For instance, helpers mises-

timate appropriate helping methods (Schroeder et al., 2017) and overestimate recipients' negative reactions after well-intentioned but unhelpful assistance (尚雪松 et al., 2021).

We therefore hypothesize that in conspicuous altruism, perspective anchoring and insufficient adjustment cause helpers to exhibit prediction bias in evaluating bystanders' assessments. Helpers base predictions on their own perspective' s known information and reference points—knowing their altruistic act and judging the social evaluation of displaying it—likely comparing conspicuous altruism to non-conspicuous altruism. This leads them to infer that others will emphasize flaunting motives and downplay altruistic motives, expecting low social evaluations. In reality, bystanders likely base evaluations on the displayed content (the altruistic act) itself, perceiving stronger altruistic motives and giving more positive evaluations than helpers predict.

1.3 Research Overview

We conducted seven studies to test prediction bias in conspicuous altruism and reveal its causes. Studies 1a, 1b, and 1c used social media posting scenarios to test whether helpers underestimate bystanders' evaluations. Study 2 retested the main effect using tangible mementos to display volunteer behavior and microblogs to showcase donation materials. Study 3 created a "Reading Poetry for You" volunteer activity, finding that after actually performing altruistic acts, participants still could not accurately predict evaluations of conspicuous altruism, demonstrating stable prediction bias. Studies 4a and 4b used verbal and social media scenarios to test the mediating roles of perceived flaunting and altruistic motives. Based on G*Power calculations (Faul et al., 2007), a minimum of 64 participants per condition is required to achieve 80% statistical power for a medium effect size (Cohen' s d = 0.5); therefore, we recruited at least 64 participants per condition in all experiments. Finally, we conducted a mini meta-analysis to test the robustness of these effects.

Study 1a

2.1.1 Participants and Design

Study 1a employed a between-subjects design (role: helper vs. bystander). We recruited 200 participants through the Credamo platform, excluding one who failed the attention check (results remained significant after exclusion), leaving 199 valid participants (135 female, aged 18-57, $M = 30.01$, $SD = 8.10$). All participants volunteered and received ¥1 compensation.

2.1.2 Procedure

First, participants received a conspicuous altruism scenario: helpers were asked to "imagine posting a social media update," while bystanders were asked to

"imagine seeing such an update" describing a ¥200 donation to the "Seeing Hope" charity project.

Next, all participants provided either evaluation predictions (helpers) or actual evaluations (bystanders). Evaluation items, adapted from Johnson (2021), included two items: "(Helper: Your friends would think you)/(Bystander: You think this person) posted behavior that is morally commendable" and "(Helper: Your friends would think you)/(Bystander: You think this person) is a virtuous individual" (1 = not at all, 7 = very much so), $\alpha = 0.72$. We analyzed the mean score. Finally, participants completed demographic measures and received compensation.

2.1.3 Results and Discussion

An independent samples t-test revealed a significant difference between roles, $t(197) = -2.91$, $p = 0.004$, Cohen' s $d = -0.41$, 95% CI = [-0.70, -0.13]. Helpers' evaluation predictions ($M = 5.55$, $SD = 0.82$) were significantly lower than bystanders' actual evaluations ($M = 5.88$, $SD = 0.80$), providing initial support for our core hypothesis.

Study 1b

Since social evaluations differ for monetary versus time-based altruism (e.g., Demczuk et al., 2023), Study 1b controlled for altruism type.

2.2.1 Participants and Design

Study 1b used a mixed design with 2 (role: helper vs. bystander, between-subjects) × 4 (scenario: monetary donation, charity shoe purchase, volunteer work, dog finding, within-subjects). We recruited 147 participants through Wenjuanxing, excluding three who failed attention checks (results remained significant), leaving 144 valid participants (118 female, aged 18–51, $M = 23.79$, $SD = 5.13$). All participants volunteered and received ¥2 compensation.

2.2.2 Procedure

The procedure mirrored Study 1a, but each participant read four different scenarios: screenshots of social media posts about "donating to the Child Promise charity," "buying shoes at a charity store" (monetary altruism), "volunteering for the Swan Heart charity," and "helping find a lost dog in the community" (time-based altruism), presented in random order.

After each scenario, participants provided evaluation predictions (helpers) or actual evaluations (bystanders) using the same items as Study 1a. Evaluation items showed high consistency across scenarios (αs ≥ 0.77) and cross-scenario reliability ($\alpha = 0.87$), so we used the mean across four scenarios for analysis. After completing all scenarios, participants provided demographics and received compensation.

2.2.3 Results and Discussion

Repeated measures ANOVA revealed a significant main effect of role, $F(1, 142) = 8.30$, $p = 0.005$, $p^2 = 0.06$; a significant main effect of scenario, $F(3, 426) = 24.57$, $p < 0.001$, $p^2 = 0.15$; and a non-significant scenario × role interaction, $F(3, 426) = 0.38$, $p = 0.771$. Independent samples t-tests showed significant mean differences across scenarios, $t(142) = -2.88$, $p = 0.005$, Cohen' s $d = -0.48$, 95% CI = [-0.82, -0.14], with helpers' predictions ($M = 4.94$, $SD = 1.04$) significantly lower than bystanders' evaluations ($M = 5.43$, $SD = 0.97$). Individual scenario analyses showed consistent results (all ps < 0.037).

Study 1b replicated Study 1a, confirming that helpers underestimate social evaluations of conspicuous altruism. Studies 1a and 1b covered both monetary and time-based altruism, demonstrating effect stability. However, they did not compare conspicuous versus non-conspicuous altruism, leaving open whether helpers generally underestimate evaluations of all altruistic acts. Study 1c therefore examined both conspicuous and non-conspicuous altruism to isolate the unique effect of display. Additionally, Study 1c expanded the dependent variable beyond moral evaluation to include interpersonal evaluations and predictions of future altruistic behavior.

Study 1c

2.3.1 Participants and Design

Study 1c used a 2 (role: helper vs. bystander) × 2 (post-altruism display: display vs. no display) between-subjects design. We recruited 400 valid participants through Credamo (all passing attention and manipulation checks; 237 female, aged 18-64, $M = 29.20$, $SD = 8.99$). All participants volunteered and received ¥1 compensation.

2.3.2 Procedure

In the conspicuous altruism condition, mirroring Study 1a, helpers imagined "posting a social media update" while bystanders imagined "seeing such an update" about participating in voluntary blood donation. In the non-conspicuous altruism condition, helpers imagined "participating in blood donation and posting a private update visible only to yourself, which a colleague accidentally saw while using your phone,"while bystanders imagined "a colleague posted a private blood donation update that you accidentally saw while using their phone."

Participants then completed attention and manipulation checks: "What charity activity did you/your colleague do?" and "How did you/your colleague see this post: A. Publicly posted B. Accidentally saw private information."

Next, participants provided evaluation predictions (helpers) or actual evaluations (bystanders). Moral evaluation items included: "(Helper: Your colleague would think you)/(Bystander: You think this person) acted how morally com-

mendably" and "(Helper: Your colleague would think you)/(Bystander: You think this person) is how virtuous" (1 = not at all, 7 = very much so), α = 0.833. Bystanders also rated interpersonal evaluations (2 items: how admirable, how worthy of friendship, α = 0.790) and predicted future altruistic behavior (2 items: likelihood of participating in other charity activities, effort in future activities, α = 0.765). Helpers predicted bystanders' ratings on these dimensions. We analyzed mean scores for moral evaluation, interpersonal evaluation, and future behavior prediction.

2.3.3 Results and Discussion

A 2 × 2 ANOVA on moral evaluation revealed a non-significant main effect of display, $F(1, 396) = 0.27$, $p = 0.603$; a significant main effect of role, $F(1, 396) = 16.95$, $p < 0.001$, $\eta^2 = 0.04$; and crucially, a significant interaction, $F(1, 396) = 14.69$, $p < 0.001$, $\eta^2 = 0.04$. Post-hoc tests showed that in the conspicuous altruism condition, helpers' predictions ($M = 5.28$, $SD = 1.17$) were significantly lower than bystanders' evaluations ($M = 6.12$, $SD = 0.64$), $t(396) = -5.62$, $p < 0.001$, Cohen's $d = -0.76$, 95% CI = [-1.08, -0.51]. In the non-conspicuous condition, predictions ($M = 5.74$, $SD = 0.97$) and evaluations ($M = 5.77$, $SD = 1.32$) did not differ, $t(396) = -0.20$, $p = 0.997$, Cohen's $d = -0.03$, 95% CI = [-0.31, 0.25]. Among helpers, conspicuous versus non-conspicuous conditions differed significantly, $t(396) = -3.08$, $p = 0.012$, Cohen's $d = -0.44$, 95% CI = [-0.72, -0.16]; among bystanders, the difference was non-significant, $t(396) = 2.34$, $p = 0.090$, Cohen's $d = 0.33$, 95% CI = [0.05, 0.61].

Analyses of interpersonal evaluation showed consistent results: non-significant main effect of display, $F(1, 396) < 0.001$, $p = 0.964$; significant main effect of role, $F(1, 396) = 11.04$, $p < 0.001$, $\eta^2 = 0.03$; significant interaction, $F(1, 396) = 14.44$, $p < 0.001$, $\eta^2 = 0.03$. In the conspicuous condition, helpers' predictions ($M = 5.25$, $SD = 1.20$) were significantly lower than bystanders' evaluations ($M = 6.00$, $SD = 0.70$), $t(396) = -4.79$, $p < 0.001$, Cohen's $d = -0.68$, 95% CI = [-0.96, -0.40]. In the non-conspicuous condition, predictions ($M = 5.64$, $SD = 1.00$) and evaluations ($M = 5.63$, $SD = 1.38$) did not differ, $t(396) = 0.10$, $p > 0.999$, Cohen's $d = 0.01$, 95% CI = [-0.26, 0.29].

For future altruistic behavior prediction, main effects were non-significant (both ps > 0.293), but the interaction was significant, $F(1, 396) = 9.34$, $p = 0.002$, $\eta^2 = 0.02$. In the conspicuous condition, helpers' predictions ($M = 5.71$, $SD = 1.10$) were significantly lower than bystanders' evaluations ($M = 6.12$, $SD = 0.58$), $t(396) = -2.91$, $p = 0.020$, Cohen's $d = -0.41$, 95% CI = [-0.69, -0.13]. No other pairwise differences were significant (all ps > 0.103).

Study 1c replicated the prediction bias in conspicuous altruism, showing it occurs only when altruism is displayed, not when it remains private. This rules out the alternative explanation that helpers generally underestimate evaluations of all their altruistic acts, confirming that the bias stems specifically from the display behavior.

However, Studies 1a, 1b, and 1c all used social media posting as the display method. Study 2 therefore employed multiple scenarios with various display methods to retest the hypothesis.

Study 2

3.1 Participants and Design

Study 2 used a mixed design with 2 (role: helper vs. bystander, between-subjects) × 3 (scenario: charity marathon T-shirt, volunteer commemorative plate, book donation microblog, within-subjects). We recruited 140 participants through Credamo (all passing attention checks; 84 female, aged 19–58, M = 30.16, SD = 7.87). All participants volunteered and received ¥2 compensation.

3.2 Procedure

The procedure mirrored Studies 1a and 1b, but each participant read three scenarios: "wearing a charity marathon T-shirt," "displaying a volunteer service commemorative plate," and "posting a book donation certificate on microblog," presented in random order. After each scenario, participants provided evaluation predictions (helpers) or actual evaluations (bystanders) using the same items as Study 1a. After completing all scenarios, participants provided demographics and received compensation.

3.3 Results and Discussion

Repeated measures ANOVA revealed a significant main effect of role, $F(1, 138) = 7.30$, $p = 0.008$, $p^2 = 0.05$; non-significant main effect of scenario and scenario × role interaction (both ps > 0.461). In the charity marathon T-shirt scenario, helpers' predictions ($M = 5.46$, $SD = 1.00$) were significantly lower than bystanders' evaluations ($M = 5.79$, $SD = 0.81$), $p = 0.030$, Cohen's d = -0.37, 95% CI = [-0.71, -0.03]. In the volunteer commemorative plate scenario, the difference was marginally significant, $p = 0.067$, Cohen's d = -0.31, 95% CI = [-0.65, 0.02], with helpers' predictions ($M = 5.36$, $SD = 1.21$) lower than bystanders' evaluations ($M = 5.72$, $SD = 1.08$). In the book donation microblog scenario, helpers' predictions ($M = 5.40$, $SD = 1.14$) were significantly lower than bystanders' evaluations ($M = 5.87$, $SD = 0.79$), $p = 0.005$, Cohen's d = -0.48, 95% CI = [-0.82, -0.14]. Across all three scenarios, helpers' predictions ($M = 5.41$, $SD = 0.95$) were significantly lower than bystanders' evaluations (M = 5.80, $SD = 0.73$), $p = 0.008$, Cohen's d = -0.46, 95% CI = [-0.80, -0.12].

Study 2 replicated the hypothesis across three scenarios with different display methods (social media and tangible objects). Study 3 then used a more ecologically valid design to further test the hypothesis.

Study 3

4.1 Participants and Design

Study 3 used a single-factor between-subjects design (role: helper vs. bystander). We recruited 140 participants through Credamo (76 female, aged 19–52, M = 26.11, SD = 6.42), randomly assigned to helper (n = 70) or bystander (n = 70) conditions. All participants volunteered and received ¥1 compensation.

4.2 Procedure

Referencing real online charity activities, we created a fictional volunteer activity called "Reading Poetry for You." Helper condition participants first read an activity description explaining that the charity served visually impaired individuals through recorded poetry readings that would be compiled into an audio collection for recipients. To ensure careful reading, participants had to view this page for at least 15 seconds before proceeding.

Helper procedure: First, participants engaged in the altruistic act by reading the activity description, then recording themselves reading Hai Zi's poem "Facing the Sea, With Spring Blossoms" (approximately one minute), requiring actual effort. Next, to conceal the study purpose, helpers read that the team had created a "Volunteer Showcase" section on the activity website to display poetry works and needed volunteers to evaluate its effectiveness. To emphasize conspicuousness, the study stressed that each volunteer could choose whether to upload their recording, which would require displaying their real name. A fictional section screenshot was included. Helpers imagined "having uploaded the audio with a current visitor listening." This page also required at least 15 seconds of viewing. After reading the scenario, helpers predicted the listener's moral evaluation of them ("How virtuous would they think you are?" 1 = extremely unvirtuous, 7 = extremely virtuous).

Bystander procedure: First, participants read the volunteer description and conspicuous scenario, learning they would evaluate the "Volunteer Showcase" section. Wording was adjusted for perspective (e.g., changing "you" to "volunteer"), with both sections requiring at least 15 seconds of reading time. Next, bystanders listened to the poem recitation audio, seeing the poem title, author, and reader's name, and learning it was randomly selected from the showcase. The audio was a pre-recorded version of "Facing the Sea, With Spring Blossoms." To avoid gender bias, bystanders heard either a male or female voice version (both approximately one minute, using the gender-neutral name "Lin Yi'an"). After listening, bystanders evaluated the reader ("How virtuous do you think this person is?").

Finally, all participants provided demographics, were debriefed, and received compensation.

4.3 Results and Discussion

An independent samples t-test on moral evaluation revealed a significant difference, $t(138) = -3.14$, $p = 0.002$, Cohen's $d = -0.53$, 95% CI = [-0.82, -0.19], with helpers' predictions ($M = 4.74$, $SD = 0.86$) significantly lower than bystanders' evaluations ($M = 5.24$, $SD = 1.01$).

Study 3's ecologically valid design replicated the prediction bias between helpers' predictions and bystanders' actual evaluations. Across Studies 1a, 1b, 2, and 3, this effect consistently emerged across various conspicuous altruism contexts. Study 4 then examined the underlying mechanism.

Study 4a

5.1.1 Participants and Design

Study 4a used a single-factor between-subjects design (role: helper vs. bystander). We recruited 140 valid participants through Credamo (81 female, aged 18-57, $M = 31.18$, $SD = 7.51$). All participants volunteered and received ¥1 compensation.

5.1.2 Procedure

First, participants received a conspicuous altruism scenario: helpers imagined mentioning their participation in a Mandarin tutoring program for mountain-area children during casual conversation with a colleague, showing related videos; bystanders imagined overhearing this conversation and seeing the videos.

Next, helpers predicted bystanders' perceptions of their flaunting and altruistic motives ("To what extent would your colleague think you mentioned this out of flaunting/altruistic motives?"); bystanders reported their perceptions of the helper' s motives ("To what extent do you think this person mentioned this out of flaunting/altruistic motives?"). Flaunting motives were measured with four self-developed items (to gain others' approval, to display virtue, to enhance image, to show off good deeds; 1 = not at all, 7 = very much so), $\alpha = 0.907$. Altruistic motives were measured with four self-developed items (to get more people involved, to raise awareness, to help more children, to promote a helping culture; 1 = not at all, 7 = very much so), $\alpha = 0.929$. We analyzed mean scores for each.

Finally, all participants provided evaluation predictions (helpers) or actual evaluations (bystanders) using two items adapted from Johnson (2021): "Your/This person' s character is virtuous" and "Your/This person' s morality is noble" (1 = not at all, 7 = very much so), $\alpha = 0.72$. Participants then completed demographics and received compensation.

5.1.3 Results and Discussion

Independent samples *t*-tests showed: significant difference in moral evaluation, $t(138) = -3.15$, $p = 0.002$, Cohen' s $d = -0.53$, 95% CI = [-0.87, -0.19], with helpers' predictions ($M = 5.14$, $SD = 1.37$) lower than bystanders' evaluations ($M = 5.82$, $SD = 1.20$). No significant difference in perceived flaunting motives, $t(138) = -1.16$, $p = 0.248$, Cohen' s $d = 0.20$, 95% CI = [-0.14, 0.53], though helpers' predictions ($M = 4.35$, $SD = 1.32$) were directionally higher than bystanders' perceptions ($M = 4.07$, $SD = 1.58$). Perceived altruistic motives differed significantly, $t(138) = 2.30$, $p = 0.023$, Cohen' s $d = -0.39$, 95% CI = [-0.73, -0.05], with helpers' predictions ($M = 5.41$, $SD = 1.28$) lower than bystanders' perceptions ($M = 5.87$, $SD = 1.10$).

Correlations showed that in both helper and bystander conditions, altruistic and flaunting motives were negatively correlated, altruistic motives positively correlated with moral evaluation, and flaunting motives negatively correlated with moral evaluation.

Using role as the independent variable (helper coded as 1, bystander as 2), perceived altruistic and flaunting motives as mediators, and moral evaluation as the dependent variable, a bootstrap analysis with 5,000 resamples showed a significant indirect effect of perceived altruistic motives, 95% CI [0.05, 0.66], indirect effect = 0.34, but not for perceived flaunting motives, 95% CI [-0.17, 0.14]. These results indicate that compared to helpers' predictions, bystanders perceived stronger altruistic motives behind conspicuous altruism, leading to more positive social evaluations than helpers anticipated.

[Figure 1: see original paper]

Study 4a replicated the main effect and preliminarily tested the mechanism, finding that perceived altruistic motives mediated the effect while perceived flaunting motives did not. Study 4b used established measures adapted from Berman et al. (2015) to retest the mechanism and, consistent with Study 1c, expanded the dependent variables to include interpersonal evaluations and future behavior predictions.

Study 4b

5.2.1 Participants and Design

Study 4b used a single-factor between-subjects design (role: helper vs. bystander). We recruited 200 participants through Credamo, excluding one who reported being 2 years old (results remained significant), leaving 199 valid participants (all passing attention checks; 80 female, aged 17–69, $M = 30.26$, $SD = 8.56$). All participants volunteered and received ¥1 compensation.

5.2.2 Procedure

First, participants received a conspicuous altruism scenario: helpers imagined "posting a social media update" while bystanders imagined "seeing such an update" about participating in voluntary blood donation. Next, helpers predicted bystanders' perceptions of their flaunting and altruistic motives, while bystanders reported their perceptions of the helper' s motives. Measures were adapted from Berman et al. (2015): flaunting motives (3 items: to show off good deeds, to impress others, to gain praise; $\alpha = 0.887$) and altruistic motives (3 items: genuinely wanting to help, sincere passion for helping, sincere concern for helping; $\alpha = 0.769$).

Then, all participants provided evaluation predictions (helpers) or actual evaluations (bystanders). Moral evaluation items (2 items: "Your/This person' s behavior is morally commendable," "Your/This person' s morality is noble" ; $\alpha = 0.677$) were followed by bystanders' interpersonal evaluations (2 items: how admirable, how worthy of friendship; $\alpha = 0.647$) and future altruistic behavior predictions (3 items: likelihood of donating blood again, likelihood of participating in other charity activities, effort in future activities; $\alpha = 0.735$). Helpers predicted bystanders' ratings on these dimensions. Finally, participants completed demographics and received compensation.

5.2.3 Results and Discussion

Independent samples t-tests revealed: significant differences in moral evaluation, $t(160.49) = -3.52$, $p < 0.001$, Cohen' s $d = -0.50$, 95% CI = [-0.78, -0.21], with helpers'predictions ($M = 5.75$, $SD = 0.95$) lower than bystanders'evaluations ($M = 6.14$, $SD = 0.56$); significant differences in interpersonal evaluation, $t(155.05) = -2.19$, $p = 0.030$, Cohen' s $d = -0.31$, 95% CI = [-0.59, -0.03], with helpers' predictions ($M = 5.83$, $SD = 1.10$) lower than bystanders'evaluations ($M = 6.10$, $SD = 0.61$); and significant differences in future behavior prediction, $t(175.61) = -2.73$, $p = 0.007$, Cohen' s $d = -0.39$, 95% CI = [-0.67, -0.10], with helpers' predictions ($M = 5.74$, $SD = 0.90$) lower than bystanders' evaluations ($M = 6.04$, $SD = 0.62$).

For potential mediators, perceived altruistic motives differed significantly, $t(197) = -2.17$, $p = 0.031$, Cohen' s $d = -0.31$, 95% CI = [-0.59, -0.03], with helpers' predictions ($M = 5.79$, $SD = 0.88$) lower than bystanders' perceptions ($M = 6.03$, $SD = 0.70$). Perceived flaunting motives also differed significantly, $t(197) = 2.32$, $p = 0.021$, Cohen' s $d = 0.33$, 95% CI = [0.05, 0.61], with helpers' predictions ($M = 4.14$, $SD = 1.62$) higher than bystanders' perceptions ($M = 3.62$, $SD = 1.54$). Thus, the independent variable significantly affected both potential mediators.

Correlations showed that in both conditions, altruistic motives positively correlated with moral evaluation, interpersonal evaluation, and future behavior prediction. Flaunting motives negatively correlated with altruistic motives (significant in bystander condition, $p = 0.011$; marginally significant in helper con-

dition, $p = 0.054$) and with all evaluation types in the bystander condition (non-significant in helper condition).

Using role as the independent variable (helper coded as 1, bystander as 2), flaunting and altruistic motives as mediators, and moral evaluation as the dependent variable, bootstrap analysis with 5,000 resamples showed a significant indirect effect of perceived altruistic motives, 95% CI [0.02, 0.32], indirect effect = 0.15, but not for perceived flaunting motives, 95% CI [-0.01, 0.02]. Results were consistent using interpersonal evaluation or future behavior prediction as dependent variables.

[Figure 2: see original paper]

Study 4b replicated Study 4a, exploring the mechanism underlying prediction bias and confirming the mediating role of perceived altruistic motives. Compared to helpers' predictions, bystanders perceived stronger altruistic motives and weaker flaunting motives behind conspicuous altruism. The mediation through perceived altruistic motives was significant, while flaunting motives showed no significant mediation, suggesting that helpers' underestimation of social evaluations stems more from underestimating bystanders' perception of altruistic motives than from overestimating perceptions of flaunting motives.

6 Mini Meta-Analysis

To integrate the seven experiments and provide additional evidence for replicability, we conducted a mini meta-analysis (or single-paper meta-analysis; Goh et al., 2016). Conducting meta-analyses within a single paper is a widely adopted robustness test (e.g., 冉雅璇 et al., 2021; Zane et al., 2020) that improves estimation precision and narrows confidence intervals (Cumming, 2014). Following McShane and Böckenholt (2017) and 冉雅璇 et al. (2021), we summarized results from all seven experiments in Table 3 . Since our primary analyses used *t*-tests, Table 3 includes means, standard deviations, sample sizes, effect sizes (Cohen's *d*), and 95% confidence intervals for each study.

The mini meta-analysis indicated an overall effect of -0.55 between helper and bystander conditions, with a 95% confidence interval of [-0.66, -0.41], excluding zero and demonstrating a robust effect.

7 General Discussion

This research reveals prediction bias in conspicuous altruism, where helpers significantly underestimate bystanders' evaluations of their actions. The effect is stable across contexts: different display methods (social media, tangible mementos, verbal mention) and altruism types (monetary donations, time donations) all show consistent prediction bias. Our mechanism exploration reveals that the bias arises from differences in perceived altruistic motives between helpers and bystanders, with no significant mediation through flaunting motives. Specifically, bystanders perceive stronger altruistic motives than helpers predict, and

they are more tolerant of mixed motives in conspicuous altruism. Helpers' inaccurate predictions and imagined negative reactions constitute the primary constraint on conspicuous altruism and a key cause of the conspicuous altruism paradox.

7.1 The Conspicuous Altruism Paradox

The conspicuous altruism paradox exists at two levels. Individually, people want others to know about their good deeds but fear being seen as boastful, worrying that display invites criticism. Societally, displaying good deeds facilitates altruism' s spread and promotes a helping culture, yet the "doing good without seeking recognition" norm may inhibit conspicuous altruism.

Conspicuous altruism holds unique value for both individuals and society. Individually, after sending altruistic signals, people gain higher status (Braun Kohlová & Urban, 2020; Hardy & Van Vugt, 2006), more cooperation opportunities (Barclay & Willer, 2007; Hardy & Van Vugt, 2006), greater mating success (Arnocky et al., 2017; Bhogal et al., 2020; Farrelly & King, 2019), more resources (Barclay, 2004), increased help from others (Engelmann & Fischbacher, 2009), and greater recognition of personal success (Black et al., 2020). Thus, people have incentives to use altruism as a costly signal to gain better social evaluations, and opportunities for good reputation effectively promote altruism (Barclay, 2012), with others' presence typically increasing helping (Bradley et al., 2018). Societally, we need to spread altruistic behavior through contagion effects, where individual acts inspire broader kindness. Charitable actions often move citizens to emulate and perform altruism (Ash, 2017; Nook et al., 2016; Romani et al., 2016), but only if the acts are visible. Without active promotion, ordinary people' s private altruism remains unobserved and cannot spread.

The conspicuous altruism paradox thus carries negative consequences. If helpers predict that displaying altruism will bring criticism, they may reduce displays, hindering altruism' s spread, and may even reduce public altruism, substantively decreasing helping behavior—neither of which society desires. In other words, conspicuous altruism could enhance individual reputation and societal helping culture, but the paradox causes individuals to miss reputation-building opportunities and society to lose chances to promote altruistic diffusion.

This research aims to reveal the paradox' s underlying causes and find solutions to promote altruistic contagion. We find that helpers' predictions are inaccurate—they underestimate bystanders' actual evaluations. This bias stems from different motive perceptions across perspectives. Bystanders perceive stronger altruistic motives and are more tolerant of mixed motives than helpers anticipate. Helpers' predictions and imagined negative reactions are the primary constraints on conspicuous altruism and a major cause of the paradox. Therefore, helpers need not overly worry about imagined criticism; they should boldly showcase their good deeds.

7.2 Unique Significance of Conspicuous Altruism

Many studies have examined how situational publicity affects altruism, but few have focused on conspicuous altruism itself. Laboratory (Andreoni & Petrie, 2004; Barclay & Willer, 2007) and field studies (Karlan & McConnell, 2014; Lacetera & Macis, 2010) show that people behave more altruistically when observed (Black & Davidai, 2020). These studies compare public versus private contexts' effects on altruism decisions, examining how altruists passively respond to situational factors. In contrast, conspicuous altruism involves altruists as active disseminators, not just passive recipients of situational influence. "Conspicuous altruism" differs from "public altruism" by focusing on individuals' decisions to actively display their altruistic acts (the display content).

Because conspicuous altruism involves active display, its conspicuousness may conceal self-interested motives, making it more problematic than public altruism. Valsesia et al. (2021) found that people face self-presentation dilemmas, wanting to showcase themselves without appearing boastful. Yang and Hsee (2022) found that encouraging (versus requiring) volunteer self-promotion reduced fundraising effectiveness, suggesting people believe flaunting motives create negative impressions, leading them to strategically conceal these motives. Our research shows helpers need not strategically inhibit conspicuous altruism due to overblown concerns—bystanders actually perceive stronger altruistic motives and are more tolerant of flaunting motives. By revealing prediction bias in this special altruistic phenomenon, we extend previous public altruism research and broaden understanding of prediction bias in helping contexts.

Furthermore, because the displayed content is altruistic, conspicuous altruism fundamentally differs from other conspicuous behaviors. While displaying status and identity may bring benefits (Anderson et al., 2015; Ball & Eckel, 1996; Huberman et al., 2004; Kraus & Mendes, 2014; Nelissen & Meijers, 2011), conspicuous consumption can backfire, generating negative evaluations (Black & Davidai, 2020; Cannon & Rucker, 2019). Such consumption signals self-interest, conflicting with prosociality (Bracha & Vesterlund, 2017; Critcher & Dunning, 2011; Lin-Healy & Small, 2013; Newman & Cain, 2014), leading to perceptions of lower cooperativeness, reduced partner selection, and damaged cooperation signals (Srna et al., 2022). However, we find that in conspicuous altruism, bystanders give unexpectedly high evaluations, demonstrating the special nature of altruistic content in reducing the negative impact of conspicuousness. This aligns with research showing that displaying status through charity rather than luxury consumption is wiser (Black & Davidai, 2020; Griskevicius et al., 2010; Hardy & Van Vugt, 2006; Sexton & Sexton, 2014)—conspicuous altruism more effectively gains respect than conspicuous consumption.

Conspicuous altruism is particularly relevant today, given the prominence of the internet and social media. Increasingly, charities attempt to expand influence through social media (Paulin et al., 2014), and ordinary users' spontaneous sharing networks help charities achieve greater impact at lower cost (Saxton &

Wang, 2014). For charities, this means not only retaining existing donors but also encouraging them to spontaneously disseminate information to reach new supporters. Our examination of prediction bias in conspicuous altruism deepens understanding of barriers to such behavior, offering insights for how charities can interact with donors to foster spontaneous sharing (Waters et al., 2009) and reach potential supporters.

7.3 Limitations and Future Directions

First, although we created various conspicuous altruism scenarios based on real-life behaviors, and Study 3 referenced actual charity activities requiring participants to perform real altruistic acts, the display behavior remained imagined rather than actively chosen. Future research could examine how helpers perceive bystander evaluations after actively choosing to display altruism and how conspicuous altruism affects subsequent behavior. Additionally, real-world factors like helper-bystander relationships, social distance, social comparisons, and background information may influence the observed prediction bias, warranting further investigation.

Second, while we examined different display and altruism methods, we did not directly measure the intensity of conspicuousness and altruism. Future research could measure these dimensions across different conspicuous altruism behaviors and assess their impact on social evaluations and prediction bias, as well as examine whether the match between conspicuousness and altruism intensity affects bystander evaluations and perspective differences.

Moreover, this research focused on ordinary people' s everyday altruism. In reality, conspicuous altruism may also involve celebrities, entrepreneurs, or corporations. Future studies could examine how bystanders evaluate different display agents and whether prediction bias exists. Additionally, uncommon or extreme altruistic acts may differ from everyday altruism and warrant investigation.

Finally, we focused on prediction bias arising from different perspectives. Future research could explore how helpers can reduce concerns about negative evaluations and actively help and display. Some studies offer insights: charities could require self-promotion, providing a salient external attribution that observers will attribute conspicuous altruism to organizational policy rather than selfish image motives, thereby reducing image concerns and promoting altruism (Yang & Hsee, 2022). Offering choice options related to competing preferences can shift the salient motive from "enhancing image" to "expressing personal views/preferences" (e.g., "I prefer cats/dogs," "I prefer strawberry/chocolate flavors"), reducing social evaluation concerns and increasing willingness to express oneself through donations (Rifkin et al., 2021). Future research could further explore how to reduce helpers' concerns about negative evaluations from conspicuousness and how bystanders can leverage conspicuous altruism' s advocacy effects to promote altruistic emulation and dissemination, allowing altruism to continue and flourish through active display.

References

The references section is preserved exactly as in the original manuscript, maintaining all citations, formatting, and Chinese-language references with their original punctuation and structure.

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