Reviewed Preprint
v1 • August 18, 2026
Not revised

For correspondence:
idongavi@umn.edu

Funding: See page 67

Reviewing editor: Toby Wise, King's College London, United Kingdom



# Social Influence in Donation Behavior: The Effect of Mean, Variance, and Psychopathy

**Tao Jin[1], Shen Zhang[2], Xinyuan Yan[3], Iris Vilares[1]**

[1] Department of Psychology, University of Minnesota - Twin Cities, Minneapolis, United States • [2] Department of Neurobiology, German Primate Center, Göttingen, Germany • [3] Neurosurgery, Baylor College of Medicine, Houston, United States

## eLife Assessment

This study reports **important** findings regarding social influence on charitable donations, showing that giving is shaped by the statistical properties of others' donations in a manner that can be captured by a reinforcement learning model. The evidence for the conclusions is **solid**, although the computational modelling could be better motivated and described, the individual differences analyses could be more robust, and some design choices could be better motivated. Overall, the core effect appears robust and is supported by multiple well-designed experiments, but the conclusions that rely upon computational modelling and individual differences may require further support.

https://doi.org/10.7554/eLife.112321.1.sa4

## Abstract

Charitable giving is a cornerstone of social welfare, and others' donations can promote individual giving. However, the computational mechanisms through which social information shapes donation behavior, and how individuals differ in their responsiveness, remain poorly understood. Here, we investigated how the statistical properties of others' donations shape individual giving, and how this was affected by individual differences in socio-affective traits. Across 4 pre-registered experiments (total N = 1,491), participants made donations before and after observing sequential donations from diverse others whose mean and variability were systematically manipulated. We found that observing generous average donations robustly increased individual giving, while observing stingy donations decreased it. Moreover, the variance in social information did not modulate this shift. In parallel, exposure to others' donations consistently reduced the variability of individual donations, with stronger reductions when others' donations were more consistent (versus more variable). Computational modeling revealed that a hybrid reinforcement learning model integrating initial donations with learned predictions of others' donations captured participants' donation behaviors well. At the individual level, changes in donations following social exposure were consistently positively associated with psychopathic traits. Model-derived parameters further revealed that people with higher psychopathic traits were more susceptible to social information, assigning greater weight to observed donations when making their second decisions. These results were replicated across all experiments, despite variations in donation magnitudes, population type, variability discrepancy, and incentive structures. Together, our findings demonstrate that the statistical structure of social information - its magnitude and variability - shapes the corresponding features of individual donations, and identify psychopathy as a key factor underlying heterogeneity in social influence.

# Introduction

Charities play vital roles in addressing societal challenges, but depend heavily on individual donations to sustain their activities. Understanding what motivates people to give is therefore critical for supporting charitable efforts. At the same time, individuals differ markedly in their willingness to donate (Auten & Rudney, 1990; Marsh et al., 2014; Nakamura et al., 2025). Identifying the sources of these differences is key to promoting prosocial behaviors at scale, strengthening social cohesion, and enhancing well-being (Aknin et al., 2013; Vanags et al., 2025).

Substantial evidence shows that individuals' choices and decisions are shaped by those of others, a phenomenon referred to as social influence (Cialdini & Goldstein, 2004; Goldstein et al., 2008; Schultner et al., 2024; Toelch & Dolan, 2015; S. Zhang et al., 2025). In the domain of charitable giving, both laboratory and field studies demonstrate that individual generosity is influenced by others' prosocial behavior (Chierchia et al., 2020; Nook et al., 2016; Shang & Croson, 2009). For instance, presenting donors with information about a large prior contribution increased individual donations in a public radio campaign (Shang & Croson, 2009), and exposure to higher average donations from others similarly elevated giving (Nook et al., 2016). These findings reveal that others' choices are a powerful driver of charitable giving.

However, most prior studies focused on how people adjust prosocial behavior in response to a single individual's decision (Molleman et al., 2022; Shang & Croson, 2009), whereas studies involving multiple others typically summarize social information using an aggregate cue such as the group average or majority level (Nook et al., 2016; Panizza et al., 2021). People now encounter social information from diverse sources in increasingly interconnected environments. In real-world fundraising, such information is often presented repeatedly and derived from multiple donors, reflecting the dynamic distribution of charitable giving beyond single-source influence. Understanding how dynamic sequences of social information shape individual giving is therefore critical to identify effective ways to promote prosocial behaviors in modern society. Yet, despite the widespread availability of diverse social information, how sequential exposure to donation information from multiple donors shapes individual giving remains poorly understood. In particular, it is unclear how the average level and variability of others' donations together influence charitable decisions.

In recent years, reinforcement learning (RL) models, characterizing direct learning from the environment, have been increasingly applied to the study of social decision-making (da Silva Pinho et al., 2024; Lockwood & Klein-Flügge, 2021; Olsson et al., 2020). These models help elucidate the dynamic processes underlying learning in social contexts (Frolichs et al., 2022; Jin et al., 2023; Joiner et al., 2017), and their extensions further capture how humans learn from others by incorporating social information into value updating and decision-making (Burke et al., 2010; Charpentier et al., 2024; Najar et al., 2020; L. Zhang & Gläscher, 2020; Zhou et al., 2024). Such approaches enable the quantification of latent cognitive processes underlying social influence. In the current study, we employed RL-based modeling to disentangle the computational mechanisms underlying changes in donation behavior in the face of diverse donations from others. Importantly, these models can derive individual-specific computational phenotypes that capture mechanistic variability associated with personality and psychopathology-related traits (Liu et al., 2025; Patzelt et al., 2018; Rhoads et al., 2024).

Social information shapes charitable giving, yet individuals vary considerably in the extent to which they incorporate others' generosity into their own decisions (Chierchia et al., 2020; Díaz-Gutiérrez et al., 2024; Lee & Chung, 2022). Such individual difference is a core feature of l functioning and offers a valuable lens for understanding how people respond to others' behavior in prosocial contexts. However, little is known about the determinants of these individual differences - namely, why some individuals substantially adjust their giving toward observed others, whereas others are less responsive. Identifying the sources of heterogeneity in social information use helps clarify when and for whom social information promotes adaptive prosocial behavior, thereby informing targeted strategies to enhance prosocial behaviors (Tump et al., 2024).

Psychopathy and empathy may influence how individuals use social information in charitable contexts. Psychopathy is a personality construct characterized by impulsivity, superficial charm, being callous and antisocial behaviors (Hare, 2006). Its core traits are increasingly recognized as varying across the general population (Guay et al., 2007; Hare & Neumann, 2008). Psychopathy has been associated with reduced prosocial behaviors (Contreras-Huerta et al., 2022; Rilling et al., 2007), diminished sensitivity to outcomes affecting others (Cutler et al., 2021; Rhoads et al., 2025), and a heightened belief in environmental volatility (Atanassova et al., 2025). In contrast, empathy refers to the capacity to understand and resonate with the affective experiences of other people (Singer & Lamm, 2009), and is linked to greater prosocial behaviors (Lockwood et al., 2014), enhanced learning in contexts involving benefits to others (Lockwood et al., 2016) and heightened sensitivity to others' pain (Singer et al., 2004). Although psychopathy and empathy are often viewed as opposing socioaffective traits relevant to prosocial motivation and charitable giving (Gunschera et al., 2022; Tusche et al., 2016), little is known about how they shape individuals' susceptibility to social influence in donation behaviors.

In the current study, we developed a social influence task on charitable giving and implemented it in four pre-registered independent experiments (n1 = 356, n2 = 372, n3 = 375, n4 = 388, total $N$ = 1491). This paradigm enabled us to examine how the statistical properties of others' donations - specifically their mean and variability - shape individual donation behaviors both in terms of donation magnitude and variability. Furthermore, we applied RL-based modeling to quantify the social information learning process underlying donation behaviors. Finally, we investigated whether individual traits, namely psychopathic traits and empathy, modulate susceptibility to social information. Across all four experiments, we found that observing generous donations led to significant increases in individual giving, whereas observing stingy donations prompted reductions in donation amounts, but that the variance of observed donations did not affect this mean shift. It did, however, affect the variability. Following exposure to others' donations, the variability of individual donations consistently decreased across conditions. Notably, this reduction in variability was stronger when others' donations were more consistent than when they were more variable. The effects of both the mean and variability of others' donations generalized to novel charitable donations. Importantly, psychopathic traits were positively associated with susceptibility to others' donations, a latent variable derived from computational modeling. This positive association between psychopathy traits and social information use appeared consistently across all studies, in both model-based and model-agnostic behavior, and even generalized to a perceptual decision-making task, where participants had to guess animals on the screen rather than make donation decisions (Molleman et al., 2019). Together, these findings demonstrate that the magnitude and variability of others' giving shaped the level and variability of individual donations, respectively, and highlight psychopathy as a key factor underlying individual differences in social information use.

# Results

Participants completed a novel social influence task on charitable giving (Fig. 1). They were randomly assigned to one of four group-norm conditions: Low Mean-Low SD (LM-LSD) of observed others' donations; Low Mean-High SD (LM-HSD); High Mean-Low SD (HM-LSD); High Mean-High SD (HM-HSD). Across all four pre-registered experiments, the task consisted of two phases. In the baseline donation phase, participants chose how much they were willing to donate to each of twenty charities, providing a measure of baseline donation tendencies. Participants then entered the observation of others' donations and second donation phase. This phase comprised twenty rounds involving the same charities. In each round, participants predicted and then observed the donations of five different other donors. Upon completing these five prediction-observation trials, they made a second donation. Donation patterns of others varied across conditions in both generosity (Low vs. High Mean) and consistency (Low vs. High SD), allowing us to quantify how these features of social information influenced individual donation shifts. Measures of psychopathy and empathy were administered after the task to probe their associations with susceptibility to social information.

In Experiment 1, college students completed a hypothetical donation task with amounts ranging from \$0 and \$2 (Table 1). To examine the robustness of social influence effects across monetary scales, the donation range was increased to \$0-\$10 and \$0-\$100 in Experiment 2 and 3, respectively, with Experiment 3 additionally introducing a larger contrast between variance conditions. Given potential discrepancies between hypothetical and real giving, Experiment 4 employed an incentive-compatible design with a more diverse Prolific sample. This experiment further examined whether social information effects generalized to donations to novel charities in the absence of social information and included a perceptual social influence task to assess the domain generality of associations between personality traits and susceptibility to social influence.

**Table 1. Overview of the four experiments.**

Experiment 1-3 recruited college students and involved hypothetical donation decisions with ranges of \$0-\$2, \$0-\$10, and \$0-\$100, respectively. Experiment 4 recruited participants from the U.S. general population via Prolific and implemented an incentive-compatible design in which donations ranged from 0 to 100 points (equivalent to \$1). Across all experiments, the mean and standard deviation of observed others' donations were experimentally manipulated (Fig. S1). In Experiment 1, we manipulated the overall mean (\$0.6 vs. \$1.4) and standard deviation (0.1 vs. 0.3) of others' donations, computed by pooling donation amounts across all charities. Experiment 2 adopted the same distributional structure as Experiment 1, with all donation values scaled by a factor of 5. Experiment 3 further extended the donation range (\$0-\$100) and increased the contrast in the standard deviation of others' donations between the Low-SD and High-SD conditions. Finally, Experiment 4 implemented an incentive-compatible design with more fine-grained donation settings, ensuring that the mean of others' donations was matched across SD conditions at the charity level while their standard deviation varied (Method S3 and Fig. S2).

| Exp | Sample | Incentive | Scale | Key Manipulation |
|---|---|---|---|---|
| 1 | College Students | Hypothetical | \$0-2 | Mean and SD of Others' Donation manipulated (group level across charities) |
| 2 | College Students | Hypothetical | \$0-10 | Same distributional structure as Exp 1 (scaled * 5) |
| 3 | College Students | Hypothetical | \$0-100 | Expanded donation range;<br>Increase contrast between Low vs. HighSD |
| 4 | U.S. general population | Incentivized | 0-100pts | Charity-level means matched between Low and HighSD conditions; SD varied |

## Observing Others' Generous (Stingy) Donations Increased (Decreased) Individuals' Donation Amounts, Regardless of the Variance in Social Information

First, we examined whether individual donations were affected by the mean and standard deviation of observed others' donations. A linear mixed-effects model on individual donation amounts revealed a significant interaction between the mean of others' donations and the phase across all four experiments (Exp 1: $b$ = -0.34, $SE$ = 0.023, $t(12652)$ = -14.83, $p < 0.001$, 95% CI = [-0.39, -0.30]; Exp 2: $b$ = -1.83, $SE$ = 0.10, $t(13432)$ = -18.16, $p < 0.001$, 95% CI = [-2.04, -1.64]; Exp 3: $b$ = -18.03, $SE$ = 1.02, $t(13627)$ = -17.71, $p < 0.001$, 95% CI = [-20.03, -16.04]; Exp 4: $b$ = -15.42, $SE$ = 0.87, $t(14563)$ = -17.63, $p < 0.001$, 95% CI = [-17.13, -13.70]; Fig. 2; Table S1). More specifically, pairwise comparison revealed that participants significantly increased their donations after observing generous donations from others, whereas participants decreased their donations after observing stingy donations (Contrast of Baseline and Shift: Exp 1: Under LowMean, $\Delta$ = 0.20, $SE$ = 0.01, $z$ = 17.96, $p < 0.001$; Under HighMean, $\triangle$ = -0.16, $SE$ = 0.01, $z$ = -13.35, $p < 0.001$; Exp 2: Under LowMean, $\Delta$ = 1.03, $SE$ = 0.05, $z$ = 20.34, $p < 0.001$; Under HighMean, $\Delta$ = -0.91, $SE$ = 0.05, $z$ = -17.88, $p < 0.001$;


**Figure 1.** **Schematic overview of the experimental design.**

**(A)** The baseline donation phase. Participants were presented with twenty charities, each accompanied by a brief description. For each charity, they were asked to decide how much they were willing to donate within the specified range. These initial responses provided a baseline measure of individual giving preferences in the absence of social information. **(B)** The observation of others' donations and the second donation phase. For each charity, participants engaged in five sequential trials in which they first predicted the donation of another individual and then received feedback about the actual donation. This procedure exposed participants to a distribution of others' giving behavior. After completing all five trials for a given charity, participants made a second donation decision for that same charity, allowing assessment of changes in giving following social information. The set of charities in this phase was identical to that in the baseline phase. **(C)** The experimental manipulation. The distribution of others' donations was manipulated along two dimensions - mean level and variability (SD) - in a 2 * 2 between-subject design. Participants were randomly assigned to one of four conditions: Low Mean-Low SD (LM-LSD); Low Mean-High SD (LM-HSD); High Mean-Low SD (HM-LSD); High Mean-High SD (HM-HSD). The histogram and table display the distributions and descriptive statistics of the manipulated donation values in Experiment 1.

Exp 3: Under LowMean, Δ = 7.75, *SE* = 0.52, *z* = 15.06, $p < 0.001$; Under HighMean, Δ = -10.80, *SE* = 0.51, *z* = -21.22, $p < 0.001$; Exp 4: Under LowMean, Δ = 8.15, *SE* = 0.44, *z* = 18.50, $p < 0.001$; Under HighMean, Δ = -7.42, *SE* = 0.43, *z* = -17.20, $p < 0.001$). The preregistered mixed ANOVA on the averaged individual amount yielded convergent evidence for the same results (Table S2), and also confirmed what we had hypothesized in the preregistrations relating to the effect of the mean. Overall, this pattern demonstrated that participants adjusted their giving behaviors toward the generosity level displayed by others. As a control analysis, we focused on the baseline phase and found no significant differences between conditions (Table S3), suggesting that the subsequent donation changes were driven by exposure to others' donations rather than by baseline variability.

In contrast, evidence for an interaction between the standard deviation of observed others' donation and phase was generally absent, appearing solely (and weakly) in Experiment 3 (Exp 3: *b* = 2.09, *SE* = 1.02, *t* (13627) = 2.05, *p* = 0.04, 95% CI = [0.09, 4.09]) but not in the remaining three experiments (Exp 1: *b* = 0.008, *SE* = 0.02, *t* (12652) = 0.34, *p* = 0.7, 95% CI = [-0.04, 0.05]; Exp 2: *b* = 0.05, *SE* =0.10, *t* (13432) = 0.50, *p* = 0.62, 95% CI = [-0.15, 0.25]; Exp 4: *b* = 0.33, *SE* = 0.86, *t* (14563) = 0.38, *p* = 0.70, 95% CI = [-1.36, 2.02]). Furthermore, contrary to what we had initially hypothesized, results from the preregistered mixed-effects ANOVA do not show any significant standard deviation and phase interaction in any of the 4 experiments (Table S2). No significant three-way interactions between the mean, standard deviation of others' donations and phase were present across all experiments (Exp 1: *b* = -0.03, *SE* = 0.03, *t*(12652) = -1.07, *p = 0.29*, 95% CI = [-0.10, 0.03]; Exp 2: *b* = -0.20, *SE* = 0.14, *t*(13432) = -1.38, *p* = 0.17, 95% CI = [-0.47, 0.08]; Exp 3: *b* = -1.04, *SE* = 1.45, *t*(13627) = -0.72, *p* = 0.47, 95% CI = [-3.88, 1.79]; Exp 4: *b* = -0.30, *SE* = 1.23, *t* (14563) = -0.25, *p* = 0.81, 95% CI = [-2.72, 2.11]). Taken together, these findings indicate that individual donation amounts were robustly shaped by the average generosity of others, with little evidence for the influence of variability of others' donations.

## Effect of Variance on Variance: The Variability of Individual Donations Decreased After Observing Others' Donations, Especially When Others Showed Less Variability in Their Giving

We then investigated whether individual donation amounts became more or less variable following exposure to others' giving, and how the average magnitude and variability of others' donations modulated these changes. Donation variability was quantified as the standard deviation of individual donations within each phase. The results of a mixed ANOVA on donation standard deviation across charities revealed a significant main effect of phase for all experiments (Exp1: $F(1, 321) = 146.63$, $p < 0.001$, partial $\eta^2 = 0.31$; Exp2: $F(1, 341) = 260.99$, $p < 0.001$, partial $\eta = 0.43$; Exp3: $F(1, 346) = 144.57$, $p < 0.001$, partial $\eta^2 = 0.30$; Exp4: $F(1, 370) = 73.18$, $p < 0.001$, partial $\eta^2 = 0.17$; Fig. 3; Table S4), demonstrating that the donation variability decreased significantly after participants observed others' donations, and this effect held under all conditions, irrespective of the distribution characteristics of others' donations. Moreover, a significant interaction between the SD of others' donations and phase was observed in Experiments 2-4 (Exp2: $F(1, 341) = 14.64$, $p < 0.001$, partial $\eta^2 = 0.04$; Exp3: $F(1, 346) = 18.44$, $p < 0.001$, partial $\eta^2 = 0.05$; Exp4: $F(1, 370) = 14.55$, $p < 0.001$, partial $\eta^2 = 0.04$), while Experiment 1 showed a similar but non-significant pattern (Exp1: $F\ (1, 321) = 2.87$, $p = 0.09$, partial $\eta^2 = 0.009$). Mirroring these results, a two-way ANOVA on the change in donation standard deviation (Shift-Baseline) revealed a significant main effect of others' donation variability. Post hoc analyses further showed that the variability in individual giving declined to a greater extent after exposure to more consistent (LowSD) than more variable (HighSD) social information in Experiment 2-4 (Contrast of HighSD and LowSD: Exp2: Under LowMean, Δ = 0.21, *SE* = 0.10, *t* (341) = 2.04, *p* = 0.04; Under HighMean, Δ = 0.34, *SE* = 0.10, *t*(341) = 3.34, $p < 0.001$; Exp3: Under LowMean, Δ = 2.46, *SE* = 1.09, *t*(346) = 2.25, *p* = 0.03; Under HighMean, Δ = 4.15, *SE* = 1.08, *t*(346) = 3.83, $p < 0.001$; Exp4: Under LowMean, Δ = 2.65, *SE* = 1.01, *t* (370) = 2.62, *p* = 0.009; Under HighMean, Δ = 2.75, *SE* = 0.99, *t*(370) = 2.78, *p* = 0.006), with Experiment 1 showing a nonsignificant trend in the same direction (Exp1: Under LowMean, Δ = 0.03, *SE* = 0.02, *t* (321) = 1.32, *p* = 0.19; Under HighMean, Δ = 0.02, *SE* = 0.02, *t*(321) = 1.08, *p* = 0.28).

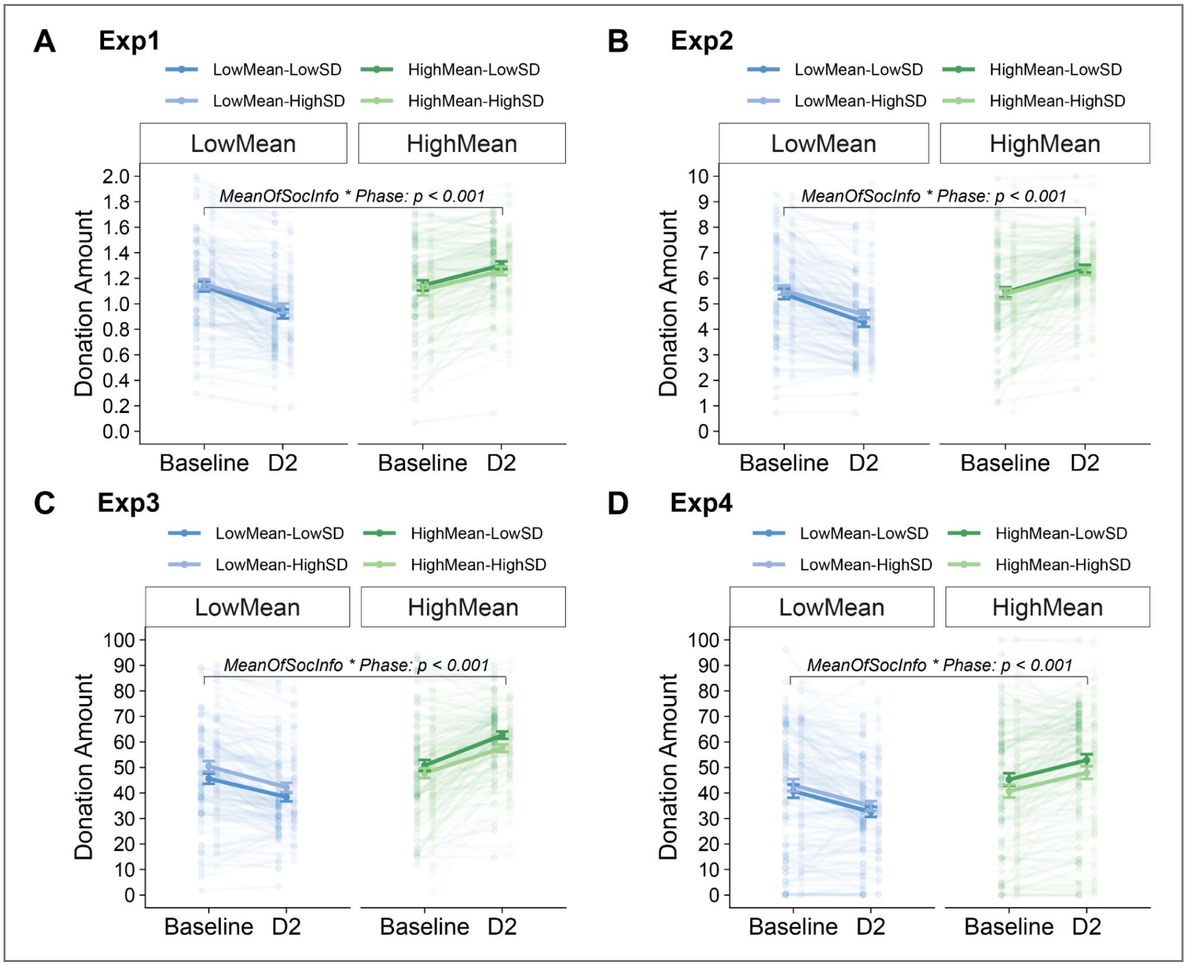



**Figure 2. The mean of observed others' donations robustly modulated shifts in individual donation amounts.**

Across all four experiments, a significant interaction between the mean of others' donations and phase was found across all four experiments (Exp 1: $b$ = -0.34, $SE$ = 0.023, $t$ (12652) = -14.83, $p$ < 0.001; Exp 2: $b$ = -1.83, $SE$ = 0.10, $t$(13432) = -18.16, $p$ < 0.001; Exp 3: $b$ = -18.03, $SE$ = 1.02, $t$(13627) = -17.71, $p$ < 0.001; Exp 4: $b$ = -15.42, $SE$ = 0.87, $t$(14563) = -17.63, $p$ < 0.001; Table S4). Specifically, participants significantly increased their own donation amounts after observing generous donations from others, but significantly decreased their donations after observing stingy donations of others. **(A - D)** displayed the results from Experiments 1 - 4. Group averages are shown as colored dots with error bars indicating ± 1 standard error of the mean (SEM), whereas faint lines and dots represent individual participants.

Thus, while the variance of social information did not seem to affect the mean donation amounts, it did influence the variance of donation amounts, such that individual donations became more consistent after exposure to more consistent others' donations than to more variable others' donations.

In Experiment 1-3, the mean and standard deviation of social information were manipulated at the aggregate level (Fig. S2). Consequently, the variability of within-charity mean donation amounts were greater in the HighSD than in the LowSD conditions, raising the possibility that the observed variability effect may partially reflect a carryover of mean-related influences. To address this concern, Experiment 4 manipulated social information at the charity-item level (See Method S3 and Fig. S2). In addition, we introduced an alternative behavioral index, termed *'pseudo-SD of individual donation'* (**See** Method S4) to quantify donation variation relative to social information rather than overall donation dispersion. This alternative analysis produced results consistent with those reported above (Result S1, Fig. S3 and Table S5). Taken together, these findings indicate that the variability cues embedded in others' donations shaped the distribution of individual givings: observing more consistent donation patterns from others narrowed the variation of individual donations, whereas the exposure to greater variability of others' donations enabled greater heterogeneity of individual givings.

## Generalization of Mean and Variability Effects of Others' Donations to Novel Charitable Giving

We next assessed whether the effect of the mean and variability of observed others' donations persisted beyond the immediate prediction-observation phase and generalized to novel charities, even in the absence of social information. In Experiment 4, after completing the observation of others' donation and second donation phase, participants made five additional donations to novel charities without receiving any further social input.

First, we focused on transfer of the mean effect. We observed a significant main effect of the mean of others' donation on the donation amount toward novel charities: people in High-Mean conditions donated more than those in Low-Mean conditions ($b$ = 16.94, $SE$ = 3.37, $t(370)$ = 5.03, $p$ < 0.001, 95% CI = [10.34, 23.54]; Table S6). All other effects were nonsignificant, including the main effect of others' donation variance and the mean and variance interaction.

Second, we investigated whether variability in observed donations modulated the variability of participants' novel giving, operationalized as the standard deviation of donations to novel charities. This analysis revealed a significant main effect of the others' donation variability ($F$(1, 370) = 9.87, $p$ = 0.002, partial $n^2$ = 0.026; Table S7). Compared with participants exposed to High-SD social information, those exposed to Low-SD social information exhibited less variability in their novel donations, with this effect being especially pronounced in the Low-Mean condition. No other effects were significant, including the main effect of others' donation mean and the mean and variance interaction. Overall, these results indicate that the statistical properties of observed donations - the mean and variability - continued to shape, respectively, the magnitude and variability of individual donations in novel contexts.

## The Hybrid Model of Initial Donations and Predictions of Others' Donations Explains Donation Behaviors Best

To formally examine trial-by-trial learning about others' donations and subsequent donation changes, we fit participants' choice behavior to a series of computational models. Specifically, the candidate model set comprised six models (See Methods for full model descriptions):

Model 0 (a non-learning model): This model serves as a baseline, assuming that both participants' predictions of others' donations and their second donations are fixed constants.

Model 1 (a D1 Only model): This model assumes that both predictions of others' donations and individual second donations are solely governed by the individual initial donations.

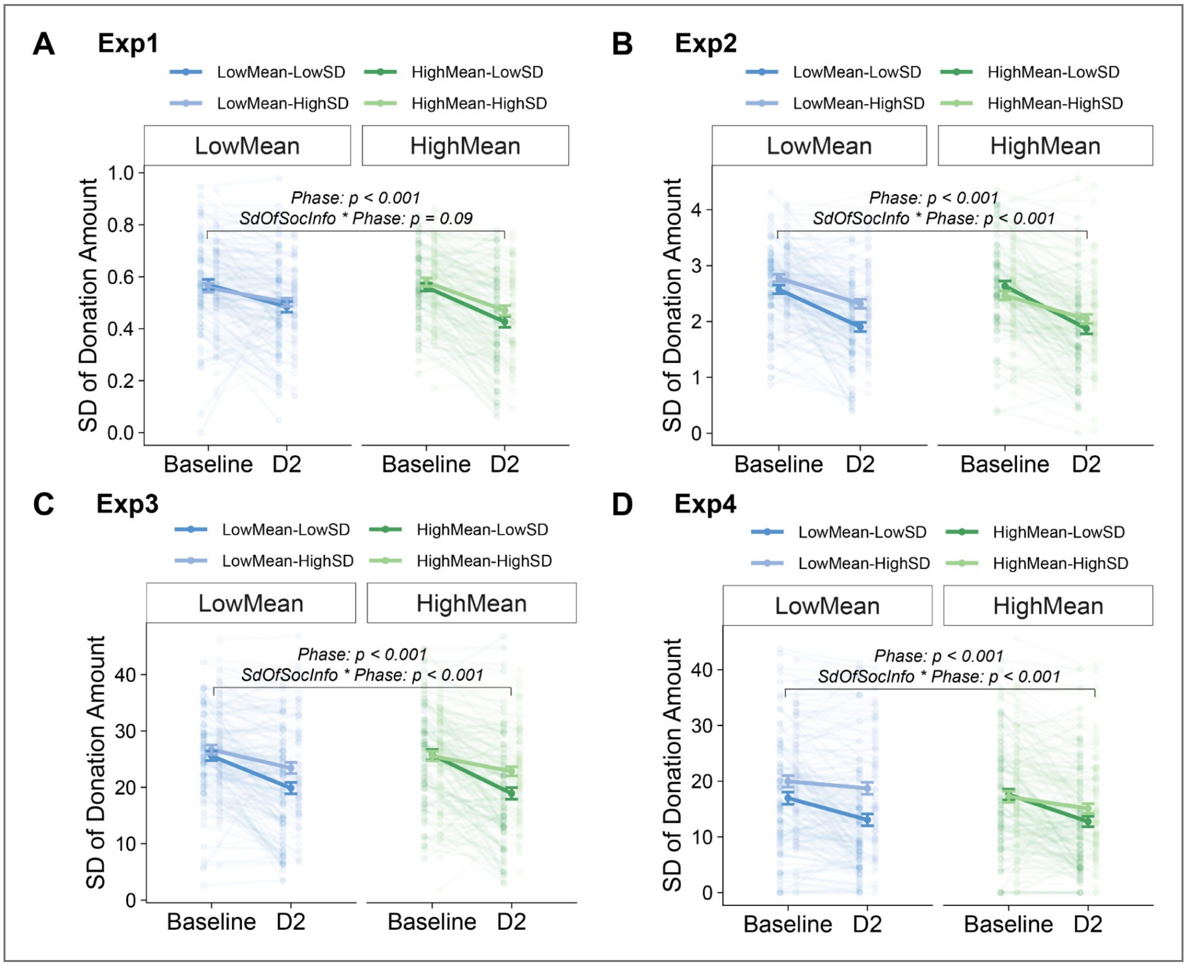



**Figure 3. Variations of individual donation amounts decreased after observing others, with a stronger reduction in Low-SD groups.**

The standard deviation of individual donation amounts significantly decreased following observing others' donations across all experiments (Exp1: $F(1, 321) = 146.63$, $p < 0.001$; Exp2: $F(1, 341) = 260.99$, $p < 0.001$; Exp3: $F(1, 346) = 144.57$, $p < 0.001$; Exp4: $F(1, 370) = 73.18$, $p < 0.001$ ; Table S4). This reduction was more pronounced when others' donations were less variable (Low-SD conditions). This pattern was significant in Experiments 2 - 4 and marginal in Experiment 1 (Exp1: $F(1, 321) = 2.87$, $p = 0.09$; Exp2: $F(1, 341) = 14.64$, $p < 0.001$ ; Exp3: $F(1, 346) = 18.44$, $p < 0.001$; Exp4: $F(1, 370) = 14.55$, $p < 0.001$). The results from Experiments 1 - 4 were displayed in **(A - D)**. Group averages are shown as colored dots with error bars indicating ± 1 standard error of the mean (SEM), whereas faint lines and dots represent individual participants.

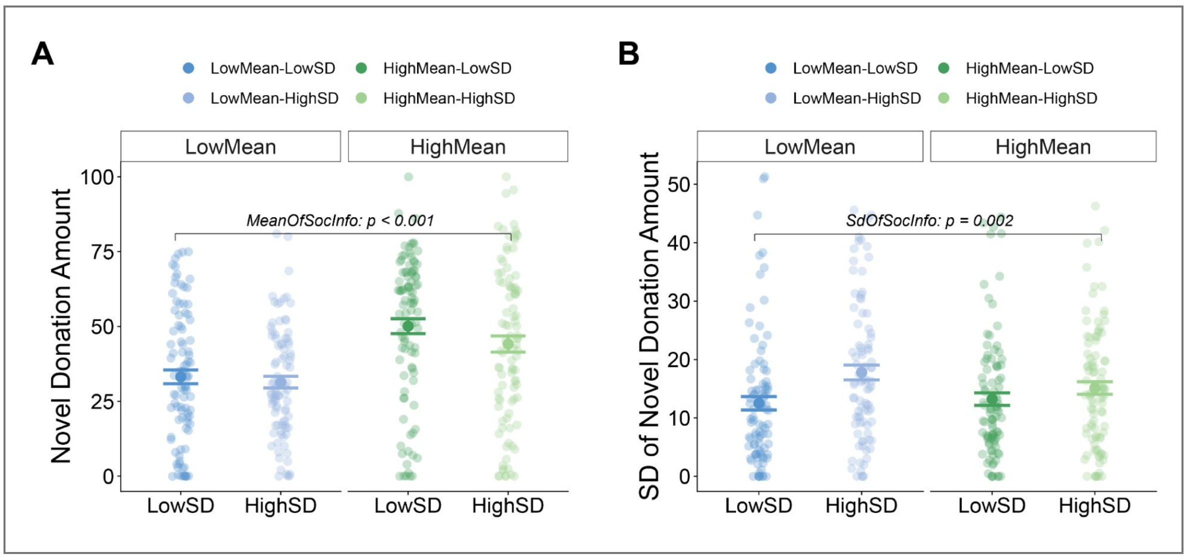



**Figure 4. The mean and variability effects of others' donations generalized to novel charitable giving.**

**(A)** After the observation of others' donation and second donation phase, participants in High-Mean conditions donated more to novel charities than those in Low-Mean conditions, despite the absence of further social input ($b$ = 16.94, $SE$ = 3.37, $t$(**370**) = 5.03, $p$ < 0.001; Table S6). **(B)** The standard deviation of individual novel donations was significantly smaller in Low-SD conditions relative to High-SD conditions, an effect most evident in Low-Mean contexts ($F$ (1,370) = 9.87, $p$ = 0.002; Table S7). Group averages are shown as colored dots with error bars indicating ± 1 standard error of the mean (SEM), whereas faint dots represent individual participants.

Model 2 (a Prediction-Only model): This model assumes that participants update their trial-by-trial predictions of others' donations using a Rescorla-Wagner rule. It also accounts for self-referential bias in the first prediction for each new charity item through a mixture of one's initial donation and the carryover prediction from the previous item. Second donations are modeled as a linear function of the updated prediction of others' donations.

Model 2b (variant of the Prediction-Only model): This model is similar to Model 2, except that it removes the self-referential anchoring assumption for the first prediction in each charity item.

Instead, the initial prediction for each new item is set to the final updated prediction from the previous item, yielding a continuous prediction sequence across items. The first prediction of the first charity item is treated as a free parameter. As in Model 2, second donations depend linearly on the updated prediction of others' donations.

Model 3 (a Hybrid model, combining the D1 and Prediction model): This model adopts the same prediction-learning process as Model 2. Second donations are modeled as a weighted combination of participants' initial donations and the updated prediction of others' donations.

Model 3b (variant of the Hybrid model, combining the D1 and Prediction model): Predictions of others' donations are updated as in Model 2b. As in Model 3a, second donations reflect a weighted average of the initial donation and the updated prediction of others' donations.

Model fittings were then compared by using the summed AIC (Akaike Information Criterion), the summed BIC (Bayesian Information Criterion) and the PXP (protected exceedance probability). Lower AIC and BIC indicate better fit after penalizing model complexity, whereas higher PXP reflects model dominance corrected for chance-level differences. Model comparison revealed that participants' behavioral data were best characterized by the Hybrid of D1 and Prediction model (Model 3) in Experiment 1 to 3, which showed the lowest AIC and BIC and highest PXP among all models (Fig. 5 and Table S8). In Experiment 4, though the variant of Hybrid of D1 and Prediction model (Model 3b) exhibited the lowest AIC and BIC, the Hybrid of D1 and Prediction model (Model 3) had a higher protected exceedance probability (PXP = 67.3%) compared to Model 3b (PXP = 32.7%), suggesting stronger evidence for Model 3 at the group level after accounting for chance. Together, these results suggest that individuals used feedback of actual donations of observed others to update their predictions of others' donation amounts, with the reliance of their own initial donation in predicting each new item diminishing over time. Importantly, the best-fitting model suggests that participants employed a hybrid strategy, combining their initial self-donation tendencies with their predictions of observed others' giving to guide their second individual donations.

We conducted model recovery and parameter recovery analyses to evaluate the model identifiability and robustness. In model recovery analyses, across all four experiments, confusion matrices were close to identity matrices, indicating that data generated by a given model was best explained by that same model (Fig. S4; See Method S5 **for model and parameter recovery details**). Additionally, to assess the robustness of the winning model, we performed parameter recovery analyses. The results showed that recovered parameters from the winning model were reliably associated with the true parameter values across all four experiments (learning rate: $r > 0.98$; relative weight on the initial donation: $r > 0.89$; social information weight: $r > 0.98$; Fig. S5).

To assess whether the winning model could adequately capture the key behavioral signatures present in the real data, we performed posterior predictive check analyses. Simulated data generated using the best-fitting parameters of each participant revealed that the winning model successfully reproduced both the mean-dependent and variance-dependent behavioral patterns. Specifically, the mean of social information modulated shifts in individual donation amounts, while the variability of social information influenced changes in the dispersion of individual donations (**See** Method S6 and Results S2, including Figs. S6-8).

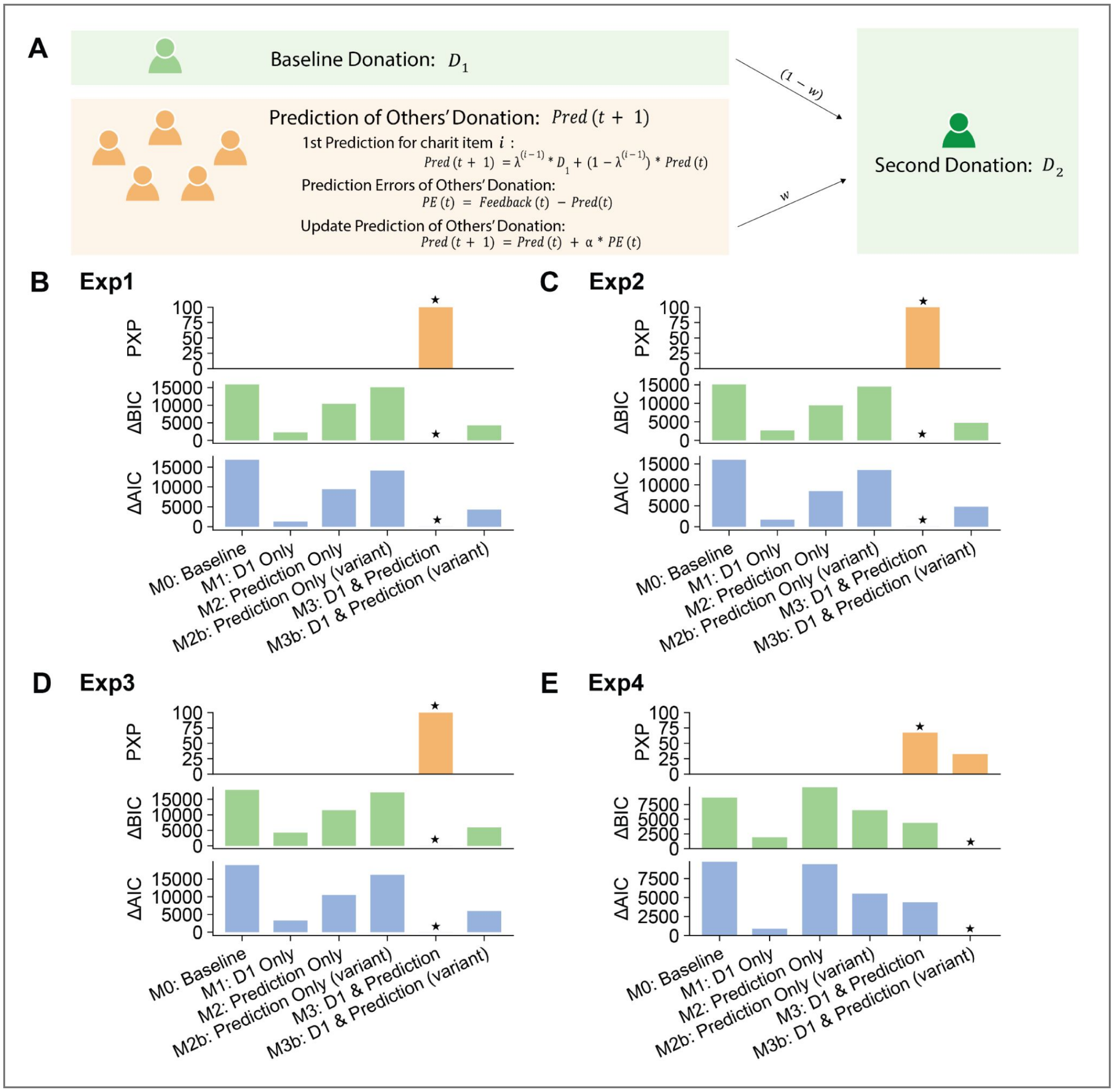



**Figure 5.** **Formal model comparison.**

**(A)** The schematic of the reinforcement learning model used to fit participants' choices. This model assumes that individuals updated their beliefs about others' donations through prediction errors, anchored to participants' initial donations as geometrically weighted references, and subsequently integrated these learned estimates with their own preferences to determine their second donations. **(B - E)** For each experiment, we computed the summed AIC (Akaike Information Criterion), the summed BIC (Bayesian Information Criterion) and the PXP (protected exceedance probability) for each candidate model (Table S8). The model with lowest summed AIC and BIC were treated as the reference model, and ΔAIC and ΔBIC values were computed by subtracting the information criterion of the reference model from that of each competing model. Smaller ΔAIC/ΔBIC values indicate better fit, and PXP values indicate stronger evidence in favor of that model at the group level. In Experiment 1 - 3, the Hybrid of D1 and Prediction model (Model 3) provided the best account of participants' behaviors, as indicated by converging evidence from all three model comparison metrics. Though a variant of the Hybrid of D1 and Prediction model (Model 3b) yielded the lowest AIC and BIC in Experiment 4, the Hybrid of D1 and Prediction model (Model 3) showed a higher PXP value, suggesting greater population-level support. Collectively, these model comparison results indicate that participants relied on both their initial donation and the updated prediction of others' donation to guide their second donation, and used feedback about others' actual donations to update their predictions. Asterisks (*) denote the best-fitting model according to each model comparison metric.

## Psychopathy Was Positively Associated with Susceptibility to Others' Donations

We subsequently assessed how individual differences in psychopathy modulated the extent to which individuals changed their donations in the light of social information. We examined it by using both model-agnostic and model-based indexes.

We first focused on absolute donation changes, capturing the magnitude of changes in individual donations after observing others' choices. The analysis revealed that absolute donation changes from the first to the second individual donations were significantly positively correlated with psychopathy scores in Experiment 2 to Experiment 4, with a marginally positive association observed in Experiment 1 (Exp1: Spearman $\rho = 0.10$, $p = 0.071$, 95% CI: [-0.01, 0.21]; Exp2: Spearman $\rho = 0.17$, $p = 0.002$, 95% CI: [0.06, 0.27]; Exp3: Spearman p = 0.14, $p = 0.009$, 95% CI: [0.03, 0.24]; Exp4: $\rho = 0.19$, $p < 0.001$, 95% CI: [0.09, 0.29]; Fig. S9). Secondly, we focused on the social information weight parameter ($w$) obtained from the winning model, Model 3 (Hybrid of D1 and Prediction Model). This parameter captured how strongly individuals incorporated predicted donations of others into their second individual donations relative to their initial donations, where higher value denoted stronger reliance on others' predicted donations. Significant positive associations between such social information weight parameter and psychopathy score were found across all four experiments (Exp1: Spearman $\rho = 0.11$, $p = 0.046$, 95% CI: [0, 0.22]; Exp2: Spearman $\rho = 0.18$, $p = 0.001$, 95% CI: [0.07, 0.28]; Exp3: Spearman $\rho = 0.14$, $p = 0.011$, 95% CI: [0.03, 0.24]; Exp4: $\rho = 0.28$, $p < 0.001$, 95% CI: [0.19, 0.38]; Fig. 6).

These results revealed a robust positive association between the susceptibility to others' donations and psychopathy. However, it remained unclear whether this relationship was independent of other factors. To address this, we conducted multiple linear regression analyses controlling for potential confounds. We first examined absolute donation changes while controlling for age, gender, and experimental conditions. The positive relationships between absolute donation changes and psychopathy remained significant across all experiments (Exp1: $b = 0.002$, $SE = 0.001$, $t$ (318) = 2.00, $p = 0.045$, 95% CI = [0.000, 0.003]; Exp2: $b = 0.011$, $SE = 0.005$, $t$ (338) = 2.28, $p = 0.023$, 95% CI = [0.001, 0.020]; Exp3: $b = 0.140$, $SE = 0.045$, $t$ (343) = 3.06, $p = 0.002$, 95% CI = [0.049, 0.225]; Exp4: $b = 0.140$, $SE = 0.040$, $t(367) = 3.56$, $p < 0.001$, 95% CI = [0.064, 0.222]; Table S9). Second, when examining the social information weight parameter, we additionally controlled other modeling parameters from the winning model along with age, gender, and experimental conditions. The social information weight remained a significant predictor of psychopathy scores in Experiment 2 to Experiment 4 with a marginal significant association in Experiment 1 (Exp1: $b = 0.002$, $SE = 0.001$, $t$ (315) = 1.70, $p = 0.090$, 95% CI = [0.000, 0.005]; Exp2: $b = 0.004$, $SE = 0.001$, $t$ (336) = 3.04, $p = 0.003$, 95% CI = [0.002, 0.007]; Exp3: $b = 0.004$, $SE = 0.002$, $t(341) = 2.43$, $p = 0.015$, 95% CI = [0.001, 0.007]; Exp4: $b = 0.007$, $SE = 0.001$, $t(365) = 5.27$, $p < 0.001$, 95% CI = [0.004, 0.010]; Table S10). Collectively, these findings revealed a consistent positive relationship between the susceptibility to others' donation and psychopathy, even after controlling for confounding factors. This pattern indicates that individuals with higher psychopathy scores were more susceptible to observed others' donations, independent of other covariates.

We also evaluated the link between the social susceptibility to others' donation and empathy. We found that there were no significant relationships between empathy and absolute donation changes across all experiments (all $p > 0.08$; **Details in** Results S3; Fig. S10; Table S11). Similarly, we didn't find significant associations between empathy and the social information weight parameter in all experiments (all $p > 0.06$; Fig. S11; Table S12). As preregistered, we also tested the relationship between psychopathy, empathy, and baseline donation levels. The baseline donation amounts were negatively correlated with psychopathy in Experiment 1-3, but this relationship was not significant (and in the opposite direction) in Experiment 4. In contrast, baseline donations were positively correlated with empathy in Experiment 1, 3, and 4, but not significant in Experiment 2 (**Details in** Results S4; Fig. S12 and Fig. S13).

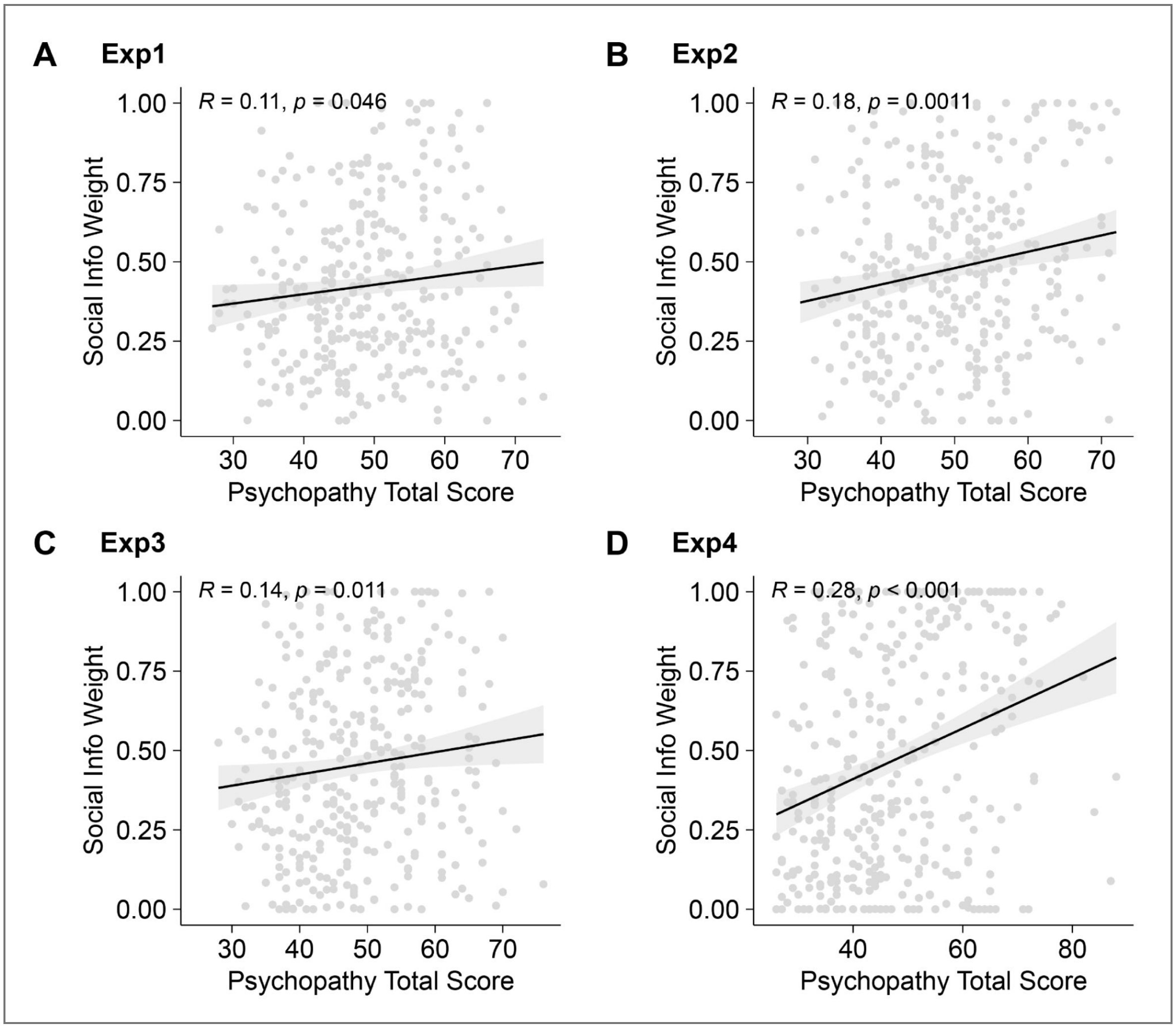



**Figure 6.** **Positive association between psychopathy and the social information weight parameter.**

The total score of psychopath was positively associated with the social information weight parameter across all experiments (Exp1: Spearman $\rho = 0.11$, $p = 0.046$; Exp2: Spearman $\rho = 0.18$, $p = 0.001$; Exp3: Spearman $\rho = 0.14$, $p = 0.011$; Exp4: $\rho = 0.28$, $p < 0.001$). **(A - D)** presented the results from Experiment 1 - 4.

### The Positive Association Between Psychopathy and Social Information Weight Generalized to Perceptual Decision-Making

A robust positive relationship between psychopathy and susceptibility to others' donations was observed in our charitable giving context. We then examine whether this association reflects a domain-specific effect or a more general tendency to incorporate social information by testing its generalization to a perceptual decision-making task. This test helps determine whether individual differences in social information use are confined to prosocial behavior or instead may reflect broader traits governing how individuals integrate social information when forming judgments. Participants in Experiment 4 additionally completed a perceptual social influence task, the BEAST (Berlin Estimate AdjuStment Task) (Molleman et al., 2019). In this task, participants made an initial estimate of the number of animals shown on the screen, then had the opportunity to revise their estimate after viewing another person's estimate. The proportional movement toward the other person's estimate was computed as the social information weight (**See** Method S7 **for more details**). We found a significant positive association between psychopathy and the social information weight in this perception social influence task (Spearman $\rho$ = 0.15, $p$ = 0.004, 95% CI: [0.04, 0.25]; Fig. 7), while psychopathy was not related to participants' initial estimates of the number of presented animals (Spearman $\rho$ = -0.03, $p$ = 0.51, 95% CI: [-0.14, 0.07]; Fig. S14). Furthermore, this positive correlation between psychopathy scores and the social information weight remained significant after accounting for gender and age ($b$ = 0.003, $SE$ = 0.001, $t(365)$ = 3.10, $p$ = 0.002, 95% CI = [0.001, 0.005]; Table S13). These findings indicate a positive relationship between psychopathy and the social information weight that generalized across both social and perceptual domains.

## Discussion

We conducted four pre-registered experiments to investigate how the magnitude and variability of others' donations influenced individual charitable giving, as well as how initial donations and the susceptibility to social information related to empathy and psychopathy. Across experiments, individuals significantly increased their personal donations after observing generous donations from others and decreased them after being exposed to stingy donations, showing that the amount others offered systematically shaped participants' own donation magnitudes. However, the variance of others' donations did not affect this mean shift. Furthermore, participants exhibited a marked reduction in donation variability after being exposed to others' donations. Individual donations converged more closely on the mean of others' donations when participants viewed more consistent, as opposed to more variable, others' donation patterns, indicating that the variability of others' donations shaped the distribution of individual giving. This social information learning process underlying donation behavior was well captured by an RL-based model, which models participants' learning about others' donations through trial-and-error updating, and then determines participants' second donations by using a weighted average of participants' initial donations and the learned social estimates. Experiment 4 further revealed that these effects of the magnitude and variation of others' donations generalized to novel charitable giving, even when no social information was provided. Furthermore, we found that the susceptibility to observed others' donations was consistently positively associated with psychopathic traits. This association also emerged in a perceptual-based social influence task, indicating a domain-general pattern across decision domains.

Across four experiments, we showed that individual donation amounts significantly increased after viewing generous donations from other donors, and decreased their donations after observing stingy others' donations. This effect remained stable regardless of changes in the maximum donation range, differences in the variability of others' donations between consistent and variable group norms, or whether the donation context was hypothetical or incentivized. Previous studies have demonstrated that others' charitable giving played an important role in shaping individual giving: individuals tended to shift their donation amounts toward the levels donated by others in both laboratory settings and field experiments (Agerström et al., 2016;

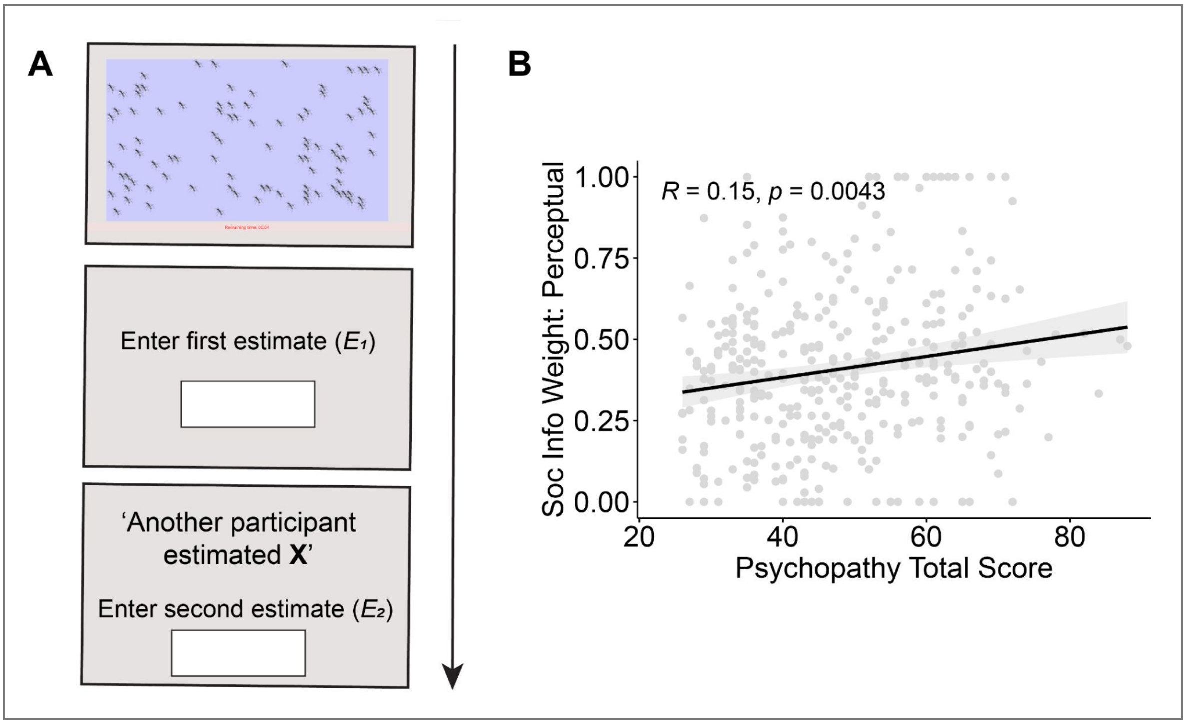



**Figure 7. Positive association between psychopathy and the social information weight in a perceptual task.**

**(A) Example trial of the BEAST Task (Berlin Estimate AdjuStment Task).** Participants were presented with an image of animals, and then estimated the number of animals displayed. They then observed another person's estimate and were given a chance to adjust their estimate. **(B)** The total psychopathy score was positively associated with the social information weight in this perceptual task (Spearman $\rho$ = 0.15, $p$ = 0.004).

Chierchia et al., 2020; Feine et al., 2023; Frey & Meier, 2004; Nook et al., 2016; Shang & Croson, 2009). Yet, these studies largely relied on a single piece of social information - either a donation from one other donor or a summary statistics representing a group of previous donors to assess how individuals respond to this single cue. Our study advances previous work by moving beyond a single social information cue to a more dynamic scenario in which individuals were sequentially exposed to diverse donations from multiple donors. Notably, even in this richer information environment, we consistently observed the same directional effects of social information, with individuals calibrating their donations toward the generosity levels of the observed donor group. In addition, evidence that the variance of others' donations influenced individual generosity levels was minimal, appearing solely (and weakly) in Experiment 3. Opportunistic conformity, characterized by the selective tendency to align with group norms when it serves individual interests, has been documented in several studies (Charness et al., 2019; Chierchia et al., 2020; Dimant, 2019). In our design, such a pattern would be reflected by larger decreases after observing stingy donations in the variable (vs. consistent) condition, and attenuated increases in donations after observing generous donations in the variable (vs. consistent) condition. However, we found limited evidence for opportunistic conformity in our study. Instead, participants tended to adjust their donations toward the donation levels of others regardless of whether observed donations were consistent or variable. Our findings thus indicate that the magnitude of observed donations might exert a more robust influence on the level of individual donations. Whether the observed donations were generous or stingy, consistent or variable, or presented in hypothetical or incentivized contexts, people reliably shifted their donations toward the observed generosity level, underscoring the magnitude of observed donations as the key determinants of changes in individual donation levels.

We observed a significant decrease in the variability of individual donations following exposure to others' donations. Notably, this effect was more pronounced when observing more consistently, compared to more variable donation patterns. One possible explanation is that this pattern may arise as a by-product of magnitude effects, such that participants shifted their donations toward the observed levels in both Low- and High-SD conditions, naturally producing corresponding differences in donation variability. We ruled out this possibility through both experimental controls and analytic approaches. In the final experiment, the mean of five others' observed donations for each item was held constant across Low- and High-SD conditions, eliminating magnitude differences as a confound. The phase-by-variability interaction persisted, despite this constraint. Additionally, we employed a novel behavioral measure, the pseudo-SD of individual donation, reflecting how closely individual donations cluster around others' mean donations. The Low-SD conditions yielded a greater reduction in pseudo-SD, demonstrating stronger alignment under consistent information. The degree to which others' opinions are homogeneous or heterogeneous influences our attitude toward their judgments (Pfänder et al., 2025), and potentially impacts personal judgments and behaviors (Brown et al., 2022; Ecker et al., 2022; Li et al., 2025; Molleman et al., 2020). Our results suggest that observing more heterogeneous charitable donations led to greater variability in individuals' own donation amounts. These findings align with accumulating evidence that low group variability was perceived as a strong and clear social signal, thereby reducing variability in individual behaviors (Brown et al., 2022; Dimant et al., 2024; Li et al., 2025). Prior work has shown that homogeneous group norms exert strong conformer pressure, pulling individuals toward the normative center, whereas heterogeneous group norms provide weaker cues and lead to more dispersed response, such as in the case of cooperation (Dimant et al., 2024). Our results extended this pattern to charitable giving and to contexts in which group norms must be learned incrementally rather than presented all at once. In our design, the social information follows normal distributions, but an important next step will be to examine whether similar effects occur in situations where observed donations follow other distribution forms. Future experimental designs that incorporate skewed or asymmetric distribution might be essential for determining how the shape and distribution of social information impact individual behaviors.

Higher levels of psychopathy have been linked to heightened antisocial behaviors and attenuated prosocial behaviors (Blair et al., 2006; Marsh, 2019; Rhoads et al., 2025). In line with these prior findings, our initial donation results revealed an inconsistent but mainly negative correlation between psychopathy and the initial amount individuals donated to charities in the absence of social cues, whereas psychopathy was unrelated to participants' initial estimates of the number of animals. Overall, these findings converge with prior evidence suggesting that individuals in higher psychopathic traits tend to be less intrinsically altruistic (Driessen et al., 2021; Gong et al., 2019; Gunschera et al., 2022; Sakai et al., 2019). Given the lower prosociality, we had initially expected that participants higher in psychopathic traits would be less susceptible to social information. Instead, we found that self-reported psychopathic traits were positively associated with the extent to which individuals incorporated social information into their donation decisions. Notably, this positive relationship remained stable across hypothetical and incentivized settings and was observed in both college student and general population samples. Previous research has suggested that people with high psychopathy were more susceptible to peers' opinions to behaviors, from evaluations of facial attractiveness, institutional misconduct to drug uses (Curtis et al., 2020; Overgaauw et al., 2019; Tatar II et al., 2016). Individuals with elevated psychopathy were grandiose, interpersonally manipulative and superficially charming. People exhibiting psychopathic traits may strategically display adaptive behaviors in response to cues during interpersonal interactions, often to maximize self-interests or achieve desired social goals (Brazil et al., 2025; Gervais et al., 2013; Osumi & Ohira, 2017). Consistent with these findings, our results might suggest that individuals with higher psychopathic traits strategically aligned with observed others' donations in order to blend in socially, thereby avoiding standing out, facilitating manipulations and strengthening their sense of fitting into the group (Brazil et al., 2023; Glenn et al., 2017; Reale et al., 2020). By adjusting individual donations toward those observed from others, they may sustain a favorable social image while still pursuing their own self-interests in the absence of explicit social cues. Such superficial compliance with social norms is consistent with how individuals high in psychopathic traits display socially desirable actions solely as a strategic tactic rather than from genuine motivations (Gunschera et al., 2022; Lin et al., 2025; Osumi & Ohira, 2017). Future studies could examine the neural mechanisms underlying social influence in individuals with elevated psychopathic traits, as identifying the key neural circuits involved would provide a more mechanistic account of why these individuals respond differently to social information. Additionally, despite the robust and replicable positive association observed between psychopathy and susceptibility to others' donations, the current study was conducted with non-clinical samples, primarily college students and individuals from the general population. Future research could investigate whether these effects extend to clinical or forensic populations, where the psychopathic traits are likely more severe.

We found that the positive link between psychopathic traits and the degree to which individuals use social cues in their decisions generalized to the perceptual domain. Our results suggest that the positive association between psychopathic traits and the susceptibility to social information emerged not only in the domain of charitable giving where individuals could infer the descriptive norm, but also in the perceptual decision-making task, where individuals merely observed others' choices. Together, these findings suggest that psychopathy may be associated with heightened domain-general susceptibility to social information - a pattern observed in both norm-driven and ambiguity-driven settings. Nevertheless, it is worth noting that the perceptual task was only done in one of the studies (study 4), and thus the perceptual task psychopathy results necessitate replication. In addition, the two paradigms employed in this study differed substantially. Future research could develop directly comparable designs across domains to better investigate the domain-specific vs. domain-general cognitive and computational mechanisms underlying social influence and how these relate to psychopathy (Lee & Chung, 2022; Olsson et al., 2020; Ruff & Fehr, 2014). Furthermore, both tasks involved two individual decisions - one made before and one after exposure to social information. This structure may have introduced demand characteristics, leading participants to adjust their response in line with perceived experimental expectations (Corneille & Lush, 2023; Iarygina et al., 2025; Orne, 1962). Individuals high in psychopathy may engage in strategic impression management, shifting their behaviors to align with these perceived expectations as a means of influencing how they were viewed by the experimenter (Gillard & Rogers, 2015; Hart et al., 2019). Future studies could explore whether the link between psychopathy and social susceptibility persists across tasks that vary in the salience of demand characteristics, from overt social cues to more subtly framed designs with immersive cover stories.

Together, we find evidence that the statistical properties of observed others' donations shaped corresponding features of individual donations: individuals increased (or decreased) their donation levels after observing generous (or stingy) levels of others' donations, while consistent patterns of observed donations yielded a stronger reduction in individual variability than variable patterns. Computational modeling successfully captured the key behavioral patterns underlying changes in charitable giving. Moreover, the susceptibility to observed donations was positively associated with psychopathic traits. These findings remained robust regardless of donation ranges, whether the discrepancy between consistent and variable observed donations was changed, whether giving was hypothetical or incentivized, or whether participants were from student or general population samples. Additionally, a positive association between social information use and psychopathic traits also appeared in a perceptual-based social influence task. These findings could have important implications for our understanding of how specific statistical features of social information systematically shape individual giving behaviors, and how individual differences, particularly in psychopathic traits, may drive variability in this process.

Our results on the mean and variability of individual donations carry direct implications for promoting charitable giving: displaying generous and consistent donor behavior may represent a simple but effective strategy for nudging people toward greater prosocial engagement. Our findings of a consistent positive relationship between psychopathic traits and susceptibility to social information also point to a practically meaningful insight: despite the lower intrinsic prosociality in individuals with elevated levels of psychopathy, presenting generous donations from others may serve as an actionable pathway for promoting prosocial behavior in this population. Overall, our results may thus directly inform the development of more effective interventions to promote charity giving at both the individual and group levels, with potential insights to promote prosocial behavior more generally.

# Methods

## Participants

Participants in Experiment 1-3 were recruited online through the course extra credit system at the University of Minnesota. Participants had to be at least 18 years old and have normal or corrected-to-normal vision. To obtain a more diverse general population sample, participants in Experiment 4 were recruited through Prolific Academic (*https://www.prolific.com*). Participation criteria included (1) United States as the country of residence, (2) age between 18 and 80 years, (3) fluency in English, and (4) normal or corrected-to-normal vision. The sample size was pre-registered, and determined based on prior work (Nook et al., 2016), which reported an effect size for the main effect of the mean of others' donations ($f = 0.23$). To account for potential publication biases, we performed an a priori power analysis in G*Power 3.1 (Faul et al., 2009) using a slightly smaller effect size ($f = 0.20$). This analysis indicated that a minimum of $N = 327$ participants was required to achieve 95% power at a significance level $\alpha = 0.05$. All participants across studies completed the experiment online. All experiments were preregistered (Exp1: *https://aspredicted.org/9jwt-48nh.pdf*; Exp2: *https://aspredicted.org/4ys2ws.pdf*; Exp3: *https://osf.io/cz2u3*; Exp4: *https://osf.io/fcugq*).

We recruited 356 participants in Experiment 1. As preregistered, 31 participants who failed more than one of seven attention checks were excluded (**See** Method S1). Data of 325 participants were thus analysed (236 females, 81 males, 8 classified themselves as non-binary, other or prefer not to respond; age = 20.24 ± 3.49 [M ± s.d.]). Experiment 2 included 372 participants. 27 participants were excluded based on the same exclusion criterion as Experiment 1, resulting in the data of 345 participants being analysed (285 females, 55 males, 5 classified themselves as non-binary, other or prefer not to respond; age = 20.05 ± 2.77 [M ± s.d.]). Experiment 3 had 375 participants, with 25 participants excluded based on the same exclusion criteria, leaving 350 participants for analysis (285 females, 56 males, 9 classified themselves as non-binary, other or prefer not to respond; age = 20.19 ± 2.78 [M ± s.d.]). Experiment 4 included 388 participants, of whom 14 were excluded according to the criteria, so 374 participants remained in the formal analysis (218 females, 151 males, 5 classified themselves as non-binary, other or prefer not to respond; age = 41.16 ± 12.73 [M ± s.d.]).

In Experiment 1-3, participants were awarded extra course credit upon completion of the experiment. In Experiment 4, participants received a fixed payment (about $10/hour), plus a bonus of up to $2 based on their decisions in the donation task and the perceptual animal counting task. One of their donations was randomly selected for actual implementation. Our study protocol has been approved by the University of Minnesota Institutional Review Board and all participants provided informed consent before participating in the study.

## Task Design - Donation Task

To investigate how the mean and variability of other's donations influenced individual givings, we employed a 2 (Mean of the observed donation: Low vs. High) x 2 (Standard Deviation (SD) of the observed donation: Low vs. High) between-subjects design. Participants were randomly assigned to one of these four group-norm conditions (e.g. the High-Mean-Low-SD of the observed donation condition). All participants completed two task phases in sequence: (1) The baseline donation phase; (2) The observation of others' donation and individual second donation phase. The experimental settings of Experiment 1 are elaborated in detail first, whereas variants across different studies are outlined subsequently. The task itself was programmed using jsPsych (de Leeuw, 2015) and run on the cognition.run platform (*https://www.cognition.run/*), and all participants performed the task online, in their own devices.

In the baseline donation phase, participants were asked to decide how much they were willing to donate to each of 20 different charities, from $0 to $2 in Experiment 1. These 20 charities were dedicated to missions in education, healthcare, environment and societal issues. The inclusion of diverse charities was used to minimize confounds from idiosyncratic charity features in measuring prosocial-influence effects (Chierchia et al., 2020). In each round, participants were first shown a brief description of a charity alongside its logo, included to enhance task engagement. To discourage uninformed decisions, the donation slider was displayed only following a 3-second presentation of charity information. Participants were then able to indicate their donations by adjusting the slider on a visual analogue scale ranging from $0 (donate nothing) to $2 (donate all). The slider's starting value was concealed until participants made their first adjustment, to avoid anchoring effects. All decisions were self-paced with no time limit. Each donation was followed by a fixation cross displayed for a randomly selected duration of 0.75, 1, or 1.25 s. The order of charities was randomized across participants. The baseline phase served to quantify individual donation propensities in the absence of any experimental manipulations.

In the observation of others' donation and individual second donation phase, participants predicted and observed donations from multiple others, and were then requested to donate again. In each round, participants were first asked to predict the first person's donation to a particular charity without a time limit. Their prediction and that person's actual donation were then displayed for at least 2 s. This prediction-then-feedback sequence was repeated for four additional others, after which participants donated to the same charity again. In total, participants completed 20 rounds (one per charity): five prediction-feedback trials per charity (100 predictions overall) followed by one individual donation per charity (20 donations total). The charity set was identical to the baseline donation phase, but was presented in a randomized order. We used jittered fixation: 0.75/ 1.0/ 1.25 s between prediction-feedback trials, and 2.25/ 2.5/ 2.75 s before the donation trial and between rounds. The first and second individual donations were probed in separate phases to allow us to quantify how individual donation changes as a function of the social-information manipulations. All presented donations from others were choices collected from real participants. Social information in Experiment 1 was obtained from a pilot study that collected real donation amounts from participants (**see** Method S2 **in SI**). In Experiment 2-4, social information was derived from both the pilot study and prior experiments with donation values rescaled to match the corresponding range of each experiment. For instance, Experiment 2 used social information from Experiment 1 and the pilot data, with others' donation amount multiplied by five to match the expended range. All participants were fully debriefed after the experiment regarding the source and nature of social information.

In Experiment 1, participants made hypothetical donations within a \$0-\$2 range (Fig. S1A). To construct the Low-Mean and High-Mean groups, we chose representing means of M = 0.6 (Low-Mean) and M = 1.4 (High-Mean), respectively, and standard deviations of SD = 0.1 (Low-SD) and SD = 0.3 (High-SD). Because we selected donation amounts based on actual participant contributions, this introduced slight differences between the conditions (Fig. S1A; Low-Mean-Low-SD condition: Mean = 0.593, SD = 0.103; Low-Mean-High-SD condition: Mean = 0.604, SD = 0.284; High-Mean-Low-SD condition: Mean = 1.400, SD = 0.115; High-Mean-High-SD condition: Mean = 1.401, SD = 0.292). Experiment 2 expanded the range to \$0-\$10 and proportionally scaled social information from Experiment 1 to preserve relative means and variances (Fig. S1B; Low-Mean-Low-SD condition: Mean = 3.00, SD = 0.514; Low-Mean-High-SD condition: Mean = 3.02, SD = 1.418; High-Mean-Low-SD condition: Mean = 7.00, SD = 0.576; High-Mean-High-SD condition: Mean = 7.00, SD =1.460). Donation remained hypothetical. In Experiment 3, the donation range was \$0-\$100, and the variability contrast between Low-SD and High-SD conditions was increased. Donations were still hypothetical (Fig. S1C; Low-Mean-Low-SD condition: Mean = 29.76, SD = 4.035; Low-Mean-High-SD condition: Mean = 30.74, SD = 16.089; High-Mean-Low-SD condition: Mean = 70.30, SD = 4.049; High-Mean-High-SD condition: Mean = 69.42, SD = 16.466). To encourage real and honest responses (Camerer & Mobbs, 2017), Experiment 4 was incentive-compatible. Donation was recorded on a 0-100-points scale (100 points = \$1) (Fig. S1D; Low-Mean-Low-SD condition: Mean = 29.65, SD = 4.191; Low-Mean-High-SD condition: Mean = 29.65, SD = 15.456; High-Mean-Low-SD condition: Mean = 69.65, SD = 4.191; High-Mean-High-SD condition: Mean = 69.65, SD = 15.456). For cleaner specification of SD effects, the per-charity mean of others' donation was matched across Low-SD and High-SD conditions (**See** Method S3).

## Procedure

In Experiment 1-3, participants first provided informed consent and read the instructions for the baseline donation phase. They then were asked to fill out demographic information (including age, gender, race, ethnicity and highest level of education). Before entering the formal task, they performed two practice donation rounds and one practice attention check round. Upon completion of the first phase, they were directed to instructions of the second phase, which involved observing others' donations and making their second donations. They then completed practice trials (two prediction trials, one individual donation trials and one attention check), before proceeding to the formal task of the second phase. At the end of the behavioral task, participants estimated the range of others' donations, and then indicated the average donation of all donors they encountered in the second phase, along with their confidence in this estimate. Afterward, they completed two questionnaires: the Levenson's Self-Report Psychopathy Scale (Levenson et al., 1995) and the Questionnaire of Cognitive and Affective Empathy (QCAE) (Reniers et al., 2011).

In Experiment 4, the procedure was the same as above with the difference that participants completed demographic information right after giving consent. Also, after completing the Levenson's Self-Report Psychopathy Scale and the QCAE, they completed a brief perceptual-based social influence task (**See** Method S7) and an additional set of questionnaires not examined here.

## Behavioral Data Analysis

To examine how the mean and variability of observed others' donations influenced individual donation tendencies, we fitted a mixed-effect linear regression, with trial-level donation amounts as the dependent variable. The independent variables included the mean of observed donation (Low vs. high), the SD of observed donation (Low vs. High), the donation phase (baseline vs. observation of others' donation and individual second donation), as well as all two-way and three-way interactions among these factors. Random intercepts were included for both participants and charity items, to account for repeated measures. The model was as formally specified as:

Donation Amount ~ MeanCond (Low vs. High) * SDCond (Low vs. High) * Phase (Baseline vs. Observation) + (1 | Item) + (1 | Subject)

We also conducted the pre-registered 2 (Mean of observed donation: Low vs. High) × 2 (SD of observed donation: Low vs. High) × 2 (Phase: Baseline vs. Shift) mixed-design ANOVA on participants' averaged donation amount.

To assess whether the average level and spread of observed others' donations modulate the variability of individual donations, we calculated the standard deviation of each participant's donations (across charities) separately for each phase. These values were then entered into a mixed-design ANOVA, with the Mean of other's donation (Low vs. High) and SD of others' donation (Low vs. High) as between-subjects factors, and Phase (Baseline vs. Observation) as a within-subjects factor.

To investigate the association between psychopathy and donation changes after exposure to others' donations, we computed Spearman correlations between the total score of Levenson's Self-Report Psychopathy Scale, QCAE and the absolute difference between participants' initial and subsequent donations.

## Computational Modeling

To quantitatively understand how individuals learned about others' donations and changed their individual donations in response to observed information, we built a series of models and compared their fit to participant's choice data.

### Model 0 (Intercept Only Model)

Model 0 assumes that participants predict other's donations using a fixed constant, and their second donation did not vary.

$$Pred\,(t+1) = b_0$$
$$D_2 = c_0$$

### Model 1 (D1 Only Model)

Model 1 assumes that participants predict other's donations and update their own second donations using the linear functions of their initial donations. It means that both predicted others' donations and second donations are modeled as regressions with the constant intercept and the slope on the initial donation.

$$Pred\,(t+1) = b_0 + b_1 {*} D_1$$
$$D_2 = c_0 + c_1 {*} D_1$$

### Model 2 (Prediction Only Model)

Model 2 employs the Rescorla-Wagner rule to model individuals' predictions of other's donations. Specifically, this model uses the prediction errors (PE) and the current estimate of others' donation to update the prediction of others' donation amount (Pred) in the next trial. The prediction error (PE) is defined as the difference between the feedback of other's actual donations and the prediction on that trial.

$$PE(t) = Feedback(t) - Pred(t)$$

The prediction of others' donation is updated based on the following equation:

$$Pred(t+1) = Pred(t) + \alpha^{*}PE(t)$$

The learning rate parameter ($\alpha$) specifies the extent to which the new information is integrated into the current beliefs (ranging between 0 and 1).

To account for the potential self-referential bias in the first prediction of each charity item, this first prediction is modeled as a mixture of the participant's initial donation and the predicted donation from the last trial of the previous charity item. Furthermore, the relative weight on the initial donation is modeled as $\lambda^{(i-1)}$, indicating it decreases geometrically with the number of charity items ($i$). At the first time ($i = 1$), the prediction is determined entirely by individual initial donations; as the item number ($i$) increases, its weight decreases, while the carryover prediction term receives an increasing complementary weight. $\lambda$ is bounded between 0 and 1. Larger $\lambda$ indicates that people place more weight on their own initial donation when forming the first prediction for a new item; a smaller $\lambda$ means predictions are primarily driven by the estimate of others' donations from the prior item.

$$Pred(t+1) = \lambda^{(i-1)*}D_1 + \left(1 - \lambda^{(i-1)}\right)^{*}Pred(t)$$

In Model 2, it assumes that the second donation depends completely on the prediction of others' donations. Specifically, the second donation is modeled as a regression on the last prediction of others' donation within each item, with the intercept and slope term.

$$D_2 = c_0 + c_1^{*}Pred(t+1)$$

### Model 2b (Prediction Only Model-Free Initial Prediction)

Model 2b mirrors Model 2, except that it omits the assumption of self-reference bias when predicting others' donations. It still allows the use of the Rescorla-Wagner rule to update predictions of others' donations. Instead, it assumes that the prediction process is modeled as a single continuous sequence: the final updated prediction from the previous item ($i$) becomes the initial prediction for the next item ($i + 1$). In addition, we assume that each person has an individual initial prediction about the other's donation. Therefore, the first prediction of others' donation is set as a free parameter (bounded between 0 and the donation range).

$$Pred(t = 1) = P_0$$

Consistent with Model 2, the second donation is specified as a linear function of the participant's prediction of others' donations, with no additional predictors.

### Model 3 (Hybrid of D1 and Prediction)

In modeling the prediction of others' donation, we adopt the same updating rule as Model 2: the prediction of others' donations is updated via the Rescorla-Wagner rule, and the first within-item prediction incorporates their initial donation geometrically over time.

For the second donation, Model 3 assumes that individuals combine both their initial donations and their prediction of observed others' donation to guide their second donations. Accordingly, the second donation is modeled as a weighted average of their initial donation and the prediction of other's donation. This weighted average is governed by a weighting parameter ($w$) bounded between 0 and 1, specifying the extent to which the second donation is based on their own initial donation ($w = 0$) versus prediction of observed donations from others ($w = 1$).

$$D_2 = w^*Pred\,(t) + (1 - w)^*D_1$$

### Model 3b (Hybrid of D1 and Prediction-Free Initial Prediction)

Like Model 2b, it assumes that updating predictions of others' donations follows the Rescorla-Wagner rule and predictions are treated as a continuous across items - the final updated prediction for one item *(i)* serves as the initial prediction for the next item ($i$ + 1).

As in Model 3, the second donation depends on both participant's initial donation for the item and the predicted donation of others via a weighted average.

## Model fitting and model comparison

All models were fitted to participants' behavioral responses to identify the model that best characterized participants' choices. Parameter optimization was conducted by using nonlinear constrained minimization (implemented in the Matlab function *fmincon*), using 500 random start points for parameters to find the best-fitting parameters for each individual. This procedure was employed to minimize the sum of squared errors (SSE) between model-predicted and actual responses across trials.

To evaluate the model performance, we used two complementary information criteria: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both criteria penalize goodness-of-fit by the model complexity. They were calculated as follows:

$$AIC = n^*\ln(\tfrac{SSE}{n}) + 2k$$
$$BIC = n^*\ln(\tfrac{SSE}{n}) + k^*\ln\,(n)$$

where n denotes the number of observations, and k represents the number of free parameters. Lower AIC and BIC indicate better model fit after accounting for model complexity. Both AIC and BIC scores were summed across participants for each model.

In addition, the posterior exceedance probability among all models was computed. This measure is used to estimate the probability that any model is more frequently used in the population than other models within the model set, while correcting for the possibility that observed differences were driven by chance rather than true population-level effects. Consistent with previous studies (Doyon et al., 2025; Frolichs et al., 2022), random-effect family-wise model comparison was conducted using the Bayesian Model Selection (BMS) procedure in the SPM12 Matlab toolbox (function *spm_BMS*) (Rigoux et al., 2014).

Model recovery and parameter recovery analyses were conducted to evaluate the identifiability and reliability of candidate models and estimated parameters. Detailed procedures are provided in the Supplementary materials (Method S5).

## Use of AI

During the manuscript preparation, ChatGPT (GPT 5.3, OpenAI) was used, solely for language polishing and wording refinement. Prompts were used for this purpose, including '*help me revise this sentence in several versions*', and '*make this sentence more concise*'. All outputs were carefully reviewed by the authors. These AI tools were not used to generate statements, design experiments, interpret findings or make scientific conclusions.

# Supplementary Materials

## Supplementary Methods

### Method S1. Attention check questions

As preregistered, seven attention check items were included for each experiment. These consisted of two catch trials in the baseline donation phase, three during the observation of others' donation and individual second donation phase, and two within the questionnaire section. The specific items are listed below.

In the baseline donation phase, attention checks (using Experiment 1 as an example; numeric ranges were proportionally changed according to the maximum donation range in Experiment 24) were:

1. This is a forced choice round: Please choose a number between \$0.6 and \$1.
2. This is a forced choice round: Please choose a number between \$1.2 and \$1.6.

In the observation of others' donation and individual donation shift phase, the following terms were used:

1. This is a forced choice round: Please choose a number between \$1.6 and \$2.
2. This is a forced choice round: Please choose a number between \$0 and \$0.4.
3. This is a forced choice round: Please choose a number between \$0.6 and \$1.

Within the questionnaires (empathy and psychopathy scales, respectively), two additional attention check questions were included:

1. Please select 'Strongly Disagree' for this question.
2. For this item please choose 'Agree somewhat'.

### Method S2. Pilot study

To obtain social information presented to participants in the social influence task on charitable giving, we conducted a pilot study with college students (N = 169; 120 females, 45 males, 4 identifying as non-binary; age = 19.98 ± 2.10 [M ± s.d.]). Participants received extra credits as compensation. In the pilot study, participants indicated their willingness to donate to a series of charities, with donation amounts ranging from \$0 to \$2. Following the donation task, participants completed the Questionnaire of Cognitive and Affective Empathy (QCAE) (Reniers et al. 2011). Donation amounts from the pilot study were presented to participants in the formal task either in their original values or after being multiplied by a constant factor. In the pilot study, individual donation amounts were significantly positively associated with empathy ($b$ = 0.007, $SE$ = 0.002, $t(146)$ = 2.80, $p$ = 0.0006, 95% CI = [0.002, 0.012]).

### Method S3. Item-level social information settings in Experimental 4

In Experiment 1-3, we systematically manipulated the mean and standard deviation of the social information across four conditions. At the aggregate level, mean donation amounts were matched between LowMean-LowSD and LowMean-HighSD conditions, as well as between HighMean-LowSD and HighMean-HighSD conditions. Likewise, overall standard deviations were matched between LowMean-LowSD and HighMean-LowSD conditions, and between LowMean-HighSD and HighMean-HighSD conditions (Fig. S1). However, these constraints were implemented at the overall level rather than at the charity-item level (Fig. S2). Consequently, the observed variability effect (participants showed lower variability in their second donation decisions compared to their first donation decisions, and more so in the lowSD conditions) may partially reflect a carryover of mean-related effects: the dispersion of the within-charity mean donation amounts tended to be greater in HighSD than in LowSD conditions. Under this structure, when participants shifted their donations toward the mean social information for each charity, the standard deviation of their donations may be expected to be larger in HighSD conditions than LowSD conditions.

To rule out this possibility, we manipulated the social information at the charity-item level in Experiment 4. Specifically, within each charity item, we held the average of the five observed donations constant between LowSD and HighSD conditions within both the LowMean and HighMean settings, while ensuring that the standard deviation of the five observed donations was higher in HighSD than in LowSD conditions. This design guarantees that the mean level of others' donations was identical at the item level across SD conditions, whereas the variability differed systematically, allowing us to test variability effects without confounding from mean differences.

## Method S4. Analysis for testing the variability effect

We further evaluated whether the variability of social information influenced changes in participants' donation variability. Importantly, instead of focusing exclusively on the variability of participants' donations per se, we assessed the extent to which participants' donation converged toward (or diverged from) the social information provided for each charity item.

To this end, we derived a behavioral index termed '*pseudo-SD of individual donation* '. This index is conceptually analogous to the conventional standard deviation, but instead of centering each participant's donation around their own mean across items, it centers individual donations around the mean donation of five observed donors for the corresponding charity item.

$$Pseudo\ SD\ of\ Individual\ Donations = \frac{\sum_{i=1}^{n}\left(Individual\ Donation_i - Mean\ of\ Five\ Others'\ Donations\right)^2}{n-1}$$

$$where\ i\ denotes\ the\ i\ th\ charity\ item;\ n = 20\ represents\ the\ total\ number\ of\ charity\ items$$

This measure provides the degree of variation in participants' donations around the mean of others' donations, with lower pseudo-SD values indicating reduced variation from others' donation levels and higher values indicating increased variation across items.

We then conducted a mixed-design ANOVA on pseudo-SD of individual donation, mirroring the analysis performed on the SD of individual donation. This analysis allows us to examine the variability effect in a way that is less susceptible to potential mean-induced differences in donation dispersion, while still characterizing the degree of variation in donation behavior.

## Method S5. Model and Parameter recovery

To evaluate whether data simulated under a given model would be best fit by the same model, we constructed confusion matrices for all four experiments. For each confusion matrix, parameter values were randomly drawn (uniformly) from predefined ranges to simulate data 360 times (85 artificial agents per condition) under each candidate model. Gaussian noise was added to the simulated data (SD = 0.1, 0.5, 5 and 5 respectively, based on the range of 2, 10, 100, and 100 in the corresponding experiment). All candidate models were subsequently fitted to the simulated data, and model recovery was assessed using BIC. Confusion matrices were based on the probabilities of selecting each candidate model as the best-fitting model for data generated by a particular model, using BIC values, with ideal recovery reflected by dominant diagonal elements.

In addition, we conducted parameter recovery analyses to evaluate whether parameters of the winning model could be reliably recovered. For each experiment, we simulated data of 360 artificial agents (85 agents per condition) using randomly sampled parameters of the winning model. Gaussian noise was added to the simulated data (SD = 0.1, 0.5, 5, and 5, respectively). The winning model was then fitted to the simulated datasets to obtain recovered parameter estimates. Parameter recovery was quantified by computing correlations between the groundtruth and recovered parameters, with higher values reflecting better recovery.

### Method S6. Posterior predictive checks

To evaluate whether the winning model (Model 3: a Hybrid model, combining the D1 and Prediction model) could reproduce the behavioral signatures present in the real data, we conducted the predictive check analyses. First, the predicted second donation amounts were generated using each participant's best-fitting parameters from the winning model. Second, the simulated data were analyzed using the same procedures applied to the empirical data. Linear mixed-effects models were fitted to donation amounts, whereas mixed-effects ANOVAs were conducted on the SD of individual donation amounts and the Pseudo SD of individual donation amounts. First donation amounts were retained from the empirical data.

### Method S7. The description of the perceptual-based social influence task

At the end of Experiment 4, participants completed a perceptual-based social influence task: the BEAST (Berlin Estimate AdjuStment Task) (Molleman, Kurvers, and van den Bos 2019). This task comprised five rounds presented in a fixed order. In each round, participants viewed an image containing 50-100 animals. After 6 seconds, the image disappeared, and participants were asked to provide an initial estimate of the number of animals *(E1 )* within 15 seconds. After entering their first estimate *(E1* ), participants were shown social information *(X*), which was an estimate provided by another person who had previously completed the task. Participants then made their second estimate *(E2* ) within 45 seconds. During the second estimate, both their first estimate and the social information were displayed on the screen.

The distance between *E1* and *X* was controlled by selecting social information from a prerecorded pool of estimates provided by 100 previous participants. Across Round 1-5, the target proportional deviation of X from *E1* (Δ)) was set to 25%, 15%, 20%, 15%, 25%, respectively. Importantly, the social information always pointed to the direction of the true value. To implement this manipulation, we computed the 'target' value (*X*') as follows: If *E1* was below the true value, *X*' = *E1* * (*1* +Δ) ; If *E1* was above the correct value, *X*' = *E1* * (*1* —Δ) ; The displayed social information was then selected as the estimate from the pre-recorded pool that was closest to the target value (*X*').

The social information use (s) in each round was calculated as a relative change toward social information: $s = (E_2 — E_1)/(X — E_1)$. Rearranging this equation yielded:$E_2 = (1 — s) * E_1 + s * E_1$, indicating that the second estimate $(E_2)$ in each round reflected a weighted average of an individual initial estimate $(E_1)$ and the social information (*X*). *s* represented a relative weight on social information.

Following previous studies (Gaule et al. 2022; Molleman et al. 2019), we excluded rounds in which participants updated their estimate in the opposite direction of the social information ($s < 0$) or moved beyond the social information ($s > 1$). We also excluded rounds in which the first or second estimate was more than 100 animals away from the true value, which likely reflected data entry errors. Rounds were also removed when either estimate was 0 or when participants' estimates exactly matched social information. In total, 92.74% of rounds were included in the final analysis.

## Supplementary Results

### Results S1. Assessing the variability effect on the dispersion of participants' donations around social information

To further probe the robustness of the variability effect and to rule out the possibility that it merely reflects mean-driven differences across social information settings, we devised an alternative behavioral index capturing the dispersion of individual donations relative to others' donations, termed the '*pseudo-SD of individual donation*' (**See** Method S4 **for details**). It parallels the conventional standard deviation, with the mean of individual donations substituted by the mean donation of five others to the specific charity. Accordingly, the pseudo-SD provides a quantification of how strongly participants' donations deviated from the social information of each charity.

A mixed ANOVA on pseudo-SD revealed robust main effects of phase across all experiments Specifically, pseudo-SD significantly decreased from baseline to observation phase, indicating that participants' donations became less dispersed around others' donation after exposure to social information (Expl: $F(1, 321) = 372.79$,$p < 0.001$, partial $\eta^2 = 0.54$; Exp2: $F(1, 341) = 464.22$, $p < 0.001$, partial $\eta = 0.58$; Exp3: $F(1, 346) = 397.08$,$p < 0.001$, partial $\eta^2 = 0.53$; Exp4: $F(1, 370) = 238.73$, $p < 0.001$, partial $\eta^2 = 0.39$; Table S5; Fig. S3). Critically, across Experiments 2-4, the interaction between the variability of others' donations and phase was significant (Exp2: $F(1, 341) = 7.82$, $p = 0.005$, partial $\eta^2 = 0.02$; Exp3: $F(1, 346) = 10.43$, $p = 0.001$, partial $\eta^2 = 0.03$; Exp4: $F(1, 370) = 10.72$, $p = 0.001$, partial $\eta^2 = 0.03$), with Experiment 1 showing a marginal trend in the same direction (Exp1: $F(1, 321) = 3.56$, $p = 0.06$, partial $\eta^2 = 0.01$). Specifically, individual donations clustered more closely around the mean of five others' donations after exposure to more consistent (LowSD) vs. more variable (HighSD) donations from others. Thus, the observed variability effect appears robust to alternative analyses and design constraints, supporting a robust role of social information variability in affecting the variability of individual giving.

## Results S2. Posterior predictive predictions

To further determine the extent to which the winning model (Model 3: a Hybrid model, combining the D1 and Prediction model) could account for the behavioral patterns observed in the empirical data, we performed the posterior predictive check analyses. We first examined the predicted donation amount, and the linear mixed-effects model applied to the simulated data showed a significant interaction between the mean of others' donations and the phase across all four experiments (Exp 1: $b = -0.29$, $SE = 0.021$, $t(12613) = -13.96$,$p < 0.001$, 95% CI = [-0.34, - 0.25]; Exp 2: $b = -1.68$, $SE = 0.09$, $t(13432) = -18.06$,$p < 0.001$, 95% CI = [-1.87, -1.50]; Exp 3: $b = -15.66$, $SE = 0.94$, $t(13627) = -16.74$,$p < 0.001$, 95% CI = [-17.49, -13.82]; Exp 4: $b = - 14.44$, $SE = 0.79$, $t(14563) = -18.27$,$p < 0.001$, 95% CI = [-16.00, -12.90]; Fig. S6). These results indicate that the effect of the average level of social information on individual donation amounts was replicated in the simulated data. We next focused on the variability effect. Mixed-effect ANOVAs on the SD of predicted individual donation amounts revealed a significant main effect of the phase across all four experiments (Expl: $F(1, 320) = 833.875$,$p < 0.001$, partial $\eta^2 = 0.72$; Exp2: $F(1, 341) = 927.77$,$p < 0.001$, partial $\eta^2 = 0.73$; Exp3: $F(1, 346) = 766.36$,$p < 0.001$, partial $\eta^2 = 0.69$; Exp4: $F(1, 370) = 443.04$,$p < 0.001$, partial $\eta^2 = 0.55$; Fig. S7). The interaction between the phase and the variability of others' donations were also reproduced in Experiments 2-4 (Exp2: $F(1, 341) = 14.64$,$p < 0.001$, partial $\eta^2 = 0.04$; Exp3: $F(1, 346) = 18.44$, $p < 0.001$, partial $\eta^2 = 0.05$; Exp4: $F(1, 370) = 14.55$, $p < 0.001$, partial $\eta^2 = 0.04$). Notably, this interaction also emerged as significant in Experiment 1 in the simulated data (Exp1: $F(1, 320) = 6.34$, $p = 0.01$, partial $\eta^2 = 0.02$). Similarly, mixed-effect ANOVAs on the Pseudo SD of predicted individual donation amounts also showed a significant main effect of the phase (Exp1: $F(1, 320) = 778.26$, p < 0.001, partial $\eta^2 = 0.71$; Exp2: $F(1, 341) = 836.19$, $p < 0.001$, partial $\eta^2 = 0.71$; Exp3: $F(1, 346) = 762.26$, $p < 0.001$, partial $\eta^2 = 0.69$; Exp4: $F(1, 370) = 480.14$, $p < 0.001$, partial $\eta^2 = 0.57$; Fig. S8). In addition, a significant interaction between the phase and the variability of others' donations was observed across all four experiments (Exp1: $F(1, 320) = 4.89$, $p = 0.03$, partial $\eta^2 = 0.02$; Exp2: $F(1, 341) = 5.08$, $p = 0.03$, partial $\eta^2 = 0.02$; Exp3: $F(1, 346) = 8.75$, $p = 0.003$, partial $\eta^2 = 0.03$; Exp4: $F(1, 370) = 7.52$, $p = 0.006$, partial $\eta^2 = 0.02$). Taken together, these posterior predictive check analyses reveal that the winning model provides a close quantitative account of the behavioral patterns observed in the empirical data.

## Results S3. Assessing the associations between empathy and the susceptibility to social information

By examining the associations between empathy and the absolute donation changes from the first to the second donation, we found there were no significant association between them in all four experiments (Exp1: Spearman $\rho = -0.04$, $p = 0.51$, 95% CI: [-0.15, 0.08]; Exp2: Spearman $\rho = - 0.03$, $p = 0.57$, 95% CI: [-0.14, 0.08]; Exp3: Spearman $\rho = -0.03$, $p = 0.58$, 95% CI: [-0.14, 0.08]; Exp4: $\rho = 0.09$, $p = 0.086$, 95% CI: [-0.02, 0.19]; Fig. S10). Similarly, there were no significant relationships between empathy and the social information weights from the model across all experiments (Exp1: Spearman $\rho = -0.01$, $p = 0.81$, 95% CI: [-0.13, 0.10]; Exp2: Spearman $\rho = - 0.10$, $p = 0.062$, 95% CI: [-0.21, 0.01]; Exp3: Spearman $\rho = -0.06$, $p = 0.24$, 95% CI: [-0.17, 0.05]; Exp4: $\rho = 0.09$, $p = 0.091$, 95%

CI: [-0.02, 0.19]; Fig. S11). After controlling for age, gender and experimental conditions, no significant effects were observed for the absolute donation changes across all experiments (Exp1: *b* = -0.000, *SE* = 0.001, *t* (318) = -0.29, *p* = 0.77, 95% CI = [-0.002, 0.001]; Exp2: *b* = -0.002, *SE* = 0.005, *t* (338) = -0.47, *p* = 0.64, 95% CI = [0.011, 0.007]; Exp3: *b* = -0.046, *SE* = 0.041, *t* (343) = -1.14, *p* = 0.25, 95% CI = [-0.126, 0.033]; Exp4: *b* = 0.039, *SE* = 0.041, *t* (367) = 0.95, *p* = 0.34, 95% CI = [-0.041, 0.119]; Table S11). Likewise, the effects of the social information weight on empathy were not significant, after additionally controlling other modeling parameters (Exp1: *b* = 0.000, *SE* = 0.001, *t* (315) = 0.02, *p* = 0.985, 95% CI = [-0.003, 0.003]; Exp2: *b* = -0.002, *SE* =0.001, *t* (336) = -1.34, *p* = 0.181, 95% CI = [-0.005, 0.001]; Exp3: *b* = -0.002, *SE* = 0.001, *t* (341) = -1.44, *p* = 0.151, 95% CI = [-0.005, 0.001]; Exp4: *b* = 0.002, *SE* = 0.001, *t*(365) = 1.47, *p* = 0.142, 95% CI = [-0.001, 0.005]; Table S12).

In addition, we investigated the association between empathy and social information use in the perceptual task, and also found no significant relationship between the two (Spearman *ρ* = 0.07, *p* = 0.18, 95% CI: [-0.03, 0.17]; Fig. S15). This effect remained null after controlling for age and gende**r** (*b* = 0.001, *SE* = 0.001, *t* (365) = 1.15, *p* = 0.251, 95% CI = [-0.001, 0.003]; Table S14).

## Results S4. Assessing the associations between psychopathy, empathy, baseline donation amount and baseline estimation

As preregistered, we also investigated the associations between psychopathy, empathy, and baseline donation levels. We had hypothesized that baseline donation amounts would be positively correlated with empathy, but negatively correlated with psychopathy. The results revealed a significant negative association between psychopathy and the baseline donation amount in Experiment 1-3, but not significant (and in the other direction) in Experiment 4 (Exp1: Spearman *ρ* = -0.15, *p* = 0.006, 95% CI: [-0.26, -0.04]; Exp2: Spearman *ρ* = -0.12, *p* = 0.024, 95% CI: [-0.23, -0.01]; Exp3: Spearman ρ = -0.11, *p* = 0.038, 95% CI: [-0.22, 0]; Exp4: *ρ* = 0.09, *p* = 0.07, 95% CI: [-0.01, 0.20]; Fig. S12). In contrast, empathy was positively associated with the baseline donation amounts in Experiment 1, 3 and 4, but not significant in Experiment 2 (Exp1: Spearman *ρ* = 0.17,*p* = 0.002, 95% CI: [0.06, 0.28]; Exp2: Spearman *ρ* = 0.02,*p* = 0.71, 95% CI: [-0.09, 0.13]; Exp3: Spearman ρ = 0.13, *p* = 0.014, 95% CI: [0.02, 0.24]; Exp4: *ρ* = 0.22,*p* < 0.001, 95% CI: [0.12, 0.32]; Fig. S13).

In the perceptual-based social influence task, and as predicted (*https://osf.io/ys9zw/overview*), psychopathy was not associated with initial estimates of the number of presented animals (Spearman ρ = -0.03, p = 0.51, 95% CI: [-0.14, 0.07]; Fig. S14). Similarly, empathy showed no significant association with participants' initial estimates (Spearman *ρ* = 0.06, *p* = 0.23, 95% CI: [-0.04, 0.17]; Fig. S16).

## Supplementary Figures

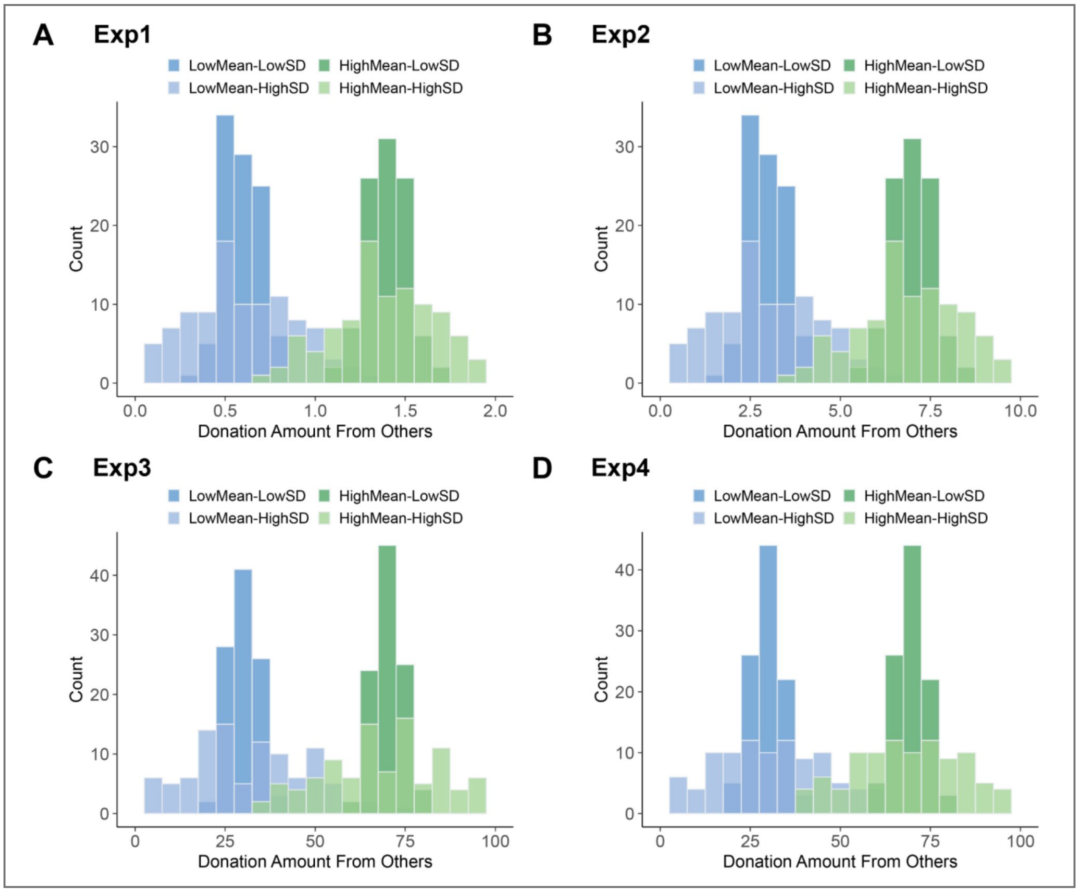



**Figure S1. The distribution of social information across conditions in all experiments.** Histograms depict the distribution of donation amounts provided by others, presented as social information to participants. In each experiment (**A - D**), participants were randomly assigned to one of four conditions in a 2 (Mean: Low vs. High) × 2 (SD: Low vs. High) between-subjects design. The distributions reflect the experimental manipulation of the mean and variability of social information in each condition (Experiment 1: LM-Low-SD condition: Mean = 0.593, SD = 0.103; Low-Mean-High-SD condition: Mean = 0.604, SD = 0.284; High-Mean-Low-SD condition: Mean = 1.400, SD = 0.115; High-Mean- High-SD condition: Mean = 1.401, SD = 0.292; Experiment 2: LSD-LSD condition: Mean = 3.00, SD = 0.514; LM-HSD condition: Mean = 3.02, SD = 1.418; HM-LSD condition: Mean = 7.00, SD = 0.576; HM-HSD condition: Mean = 7.00, SD =1.460; Experiment 3: LM-LSD condition: Mean = 29.76, SD = 4.035; LM-HSD condition: Mean = 30.74, SD = 16.089; HM-LSD condition: Mean = 70.30, SD = 4.049; HM-HSD condition: Mean = 69.42, SD = 16.466; Experiment 4: LM-LSD condition: Mean = 29.65, SD = 4.191; LM-HSD condition: Mean = 29.65, SD = 15.456; HM-LSD condition: Mean = 69.65, SD = 4.191; HM-HSD condition: Mean = 69.65, SD = 15.456).

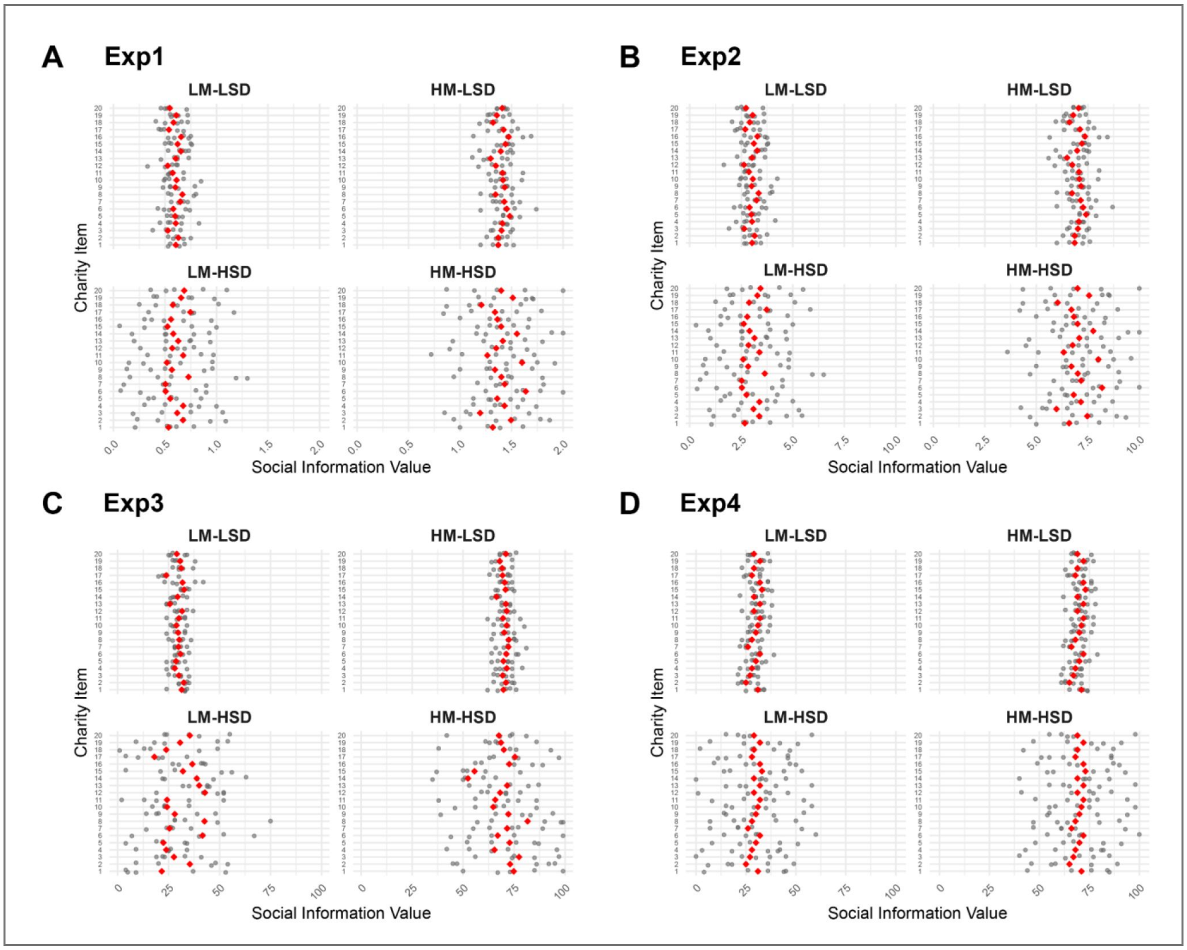



**Figure S2.** **Item-level social information values across charity items and conditions in all experiments.**

Each panel displays the observed donation amounts presented as social information for each charity item (Items 1-20) across four conditions (LowMean-LowSD, LowMean-HighSD, HighMean-LowSD, HighMean-HighSD). The x-axis represents the value of others' donation amounts, and the y-axis represents the charity item. Gray dots denote individual observed donations (five per item), and red dots represent the average of the five observed donations for each charity item. In Experiment 1 - 3 (**A - C**), the mean and standard deviation of social information were manipulated at the aggregate level across items. At the group level, HighMean conditions had larger average donation amounts than LowMean conditions, and HighSD conditions had greater variability than LowSD conditions. However, these constraints were implemented across items rather than within each item, such that item-level means differed between SD groups. In Experiment 4 (**D**), social information was manipulated at the charity-item level to eliminate this potential confound. Specifically, for each charity item, the mean of the five observed donations was held constant between LowSD and HighSD conditions, while the standard deviation of the five observed donations was systematically larger in HighSD than LowSD conditions. This design ensured that variability effects could be examined independently of the mean-related effects.

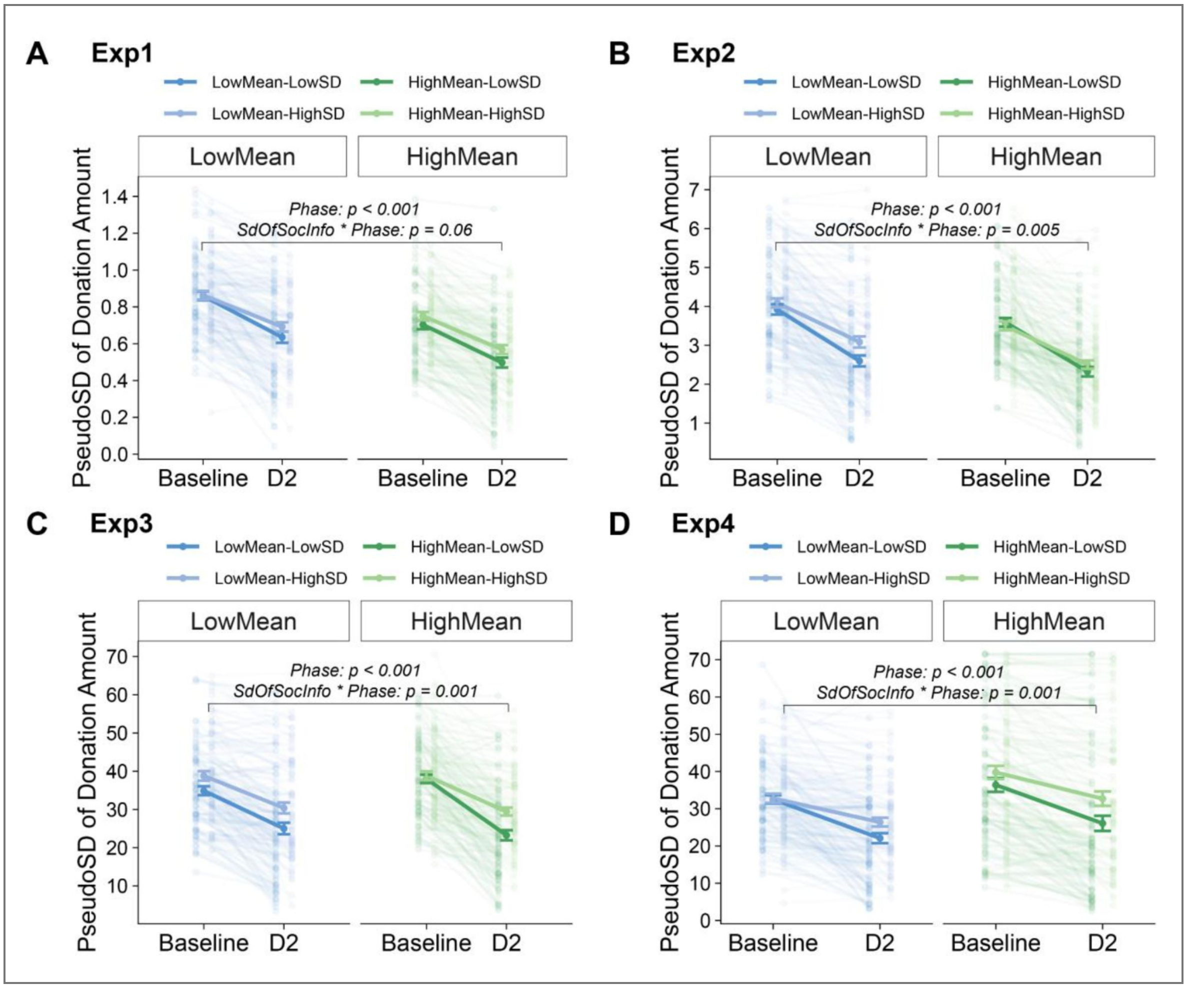



**Figure S3.** **Exposure to others' donations promoted convergence of individual donations toward group giving, especially when others' donations were consistent.**

The pseudo-SD of individual donations, indexing dispersion of individual donations around others' mean donations, significantly decreased after observing others' donations in all experiments ((Expl: $F(1, 321) = 372.79$,$p < 0.001$, partial $\eta^2 = 0.54$; Exp2: $F(1, 341) = 464.22$,$p < 0.001$, partial $\eta^2 = 0.58$; Exp3: $F(1, 346) = 397.08$,$p < 0.001$, partial $\eta^2 = 0.53$; Exp4: $F(1, 370) = 238.73$,$p < 0.001$, partial $\eta^2 = 0.39$; Table S5). This convergence effect was significantly larger in Low-SD conditions in Experiment 2 - 4 and marginal in Experiment 1 (Exp1: $F(1, 321) = 3.56$, $p = 0.06$, partial $\eta^2 = 0.01$; Exp2: $F(1, 341) = 7.82$, $p = 0.005$, partial $\eta^2 = 0.02$; Exp3: $F(1, 346) = 10.43$, $p = 0.001$, partial $\eta^2 = 0.03$; Exp4: $F(1, 370) = 10.72$, $p = 0.001$, partial $\eta^2 = 0.03$). Group averages are shown as colored dots with error bars indicating ± 1 standard error of the mean (SEM), whereas faint lines and dots represent individual participants.

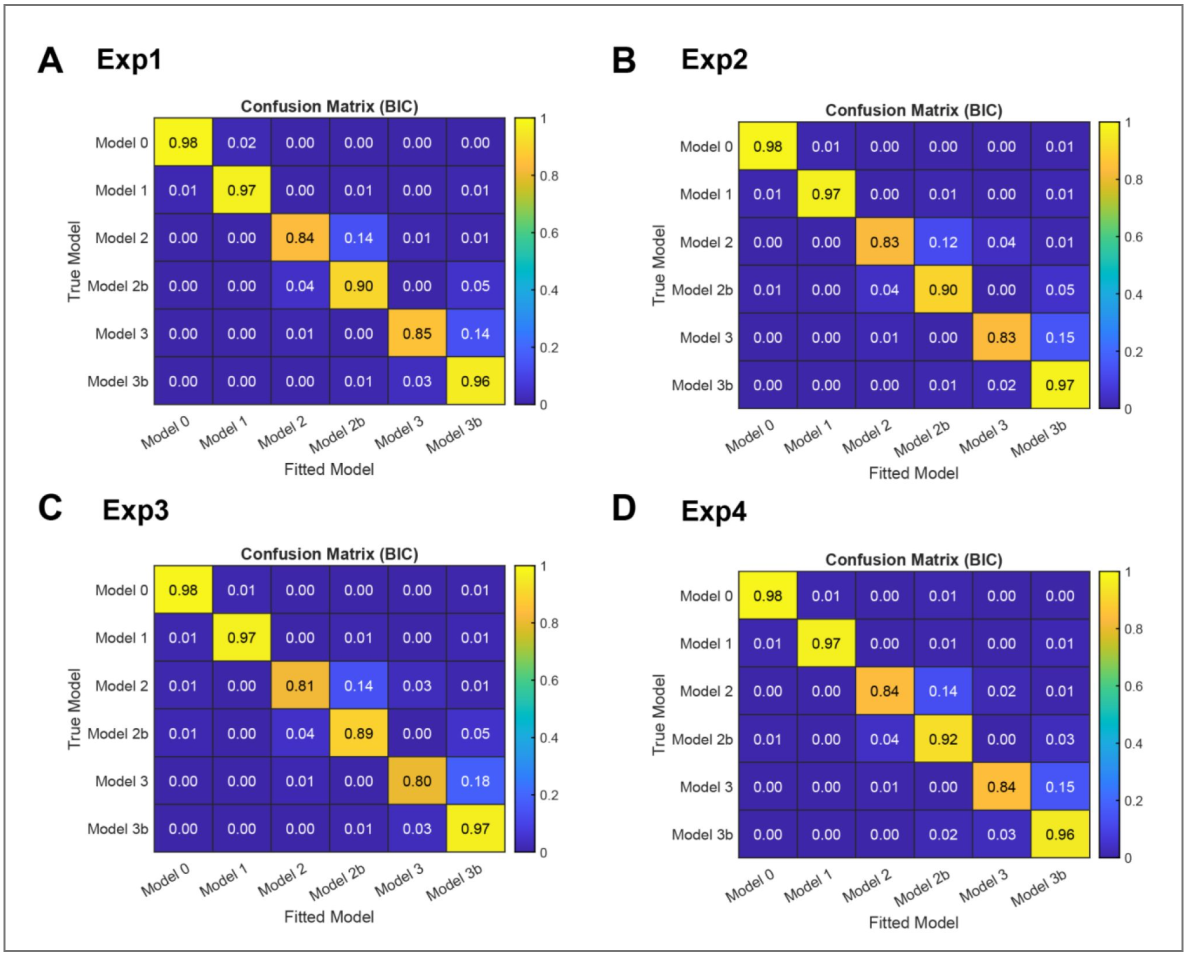



**Figure S4.** **Model recovery.**

To assess whether our tested models could be reliably distinguished, model recovery analyses were performed (**A - D**). For each experiment, simulated datasets were generated from each candidate model (N = 360 artificial agents; 85 simulated agents per condition), and Gaussian noise was added. All candidate models were fitted to the simulated data and compared. Values in the confusion matrices reflect the probabilities that a candidate model is selected as the best-fitting model for data generated from a given model, according to BIC. Across all experiments, the confusion matrices exhibited strong diagonal dominance, demonstrating great model recovery and clear separation among models.

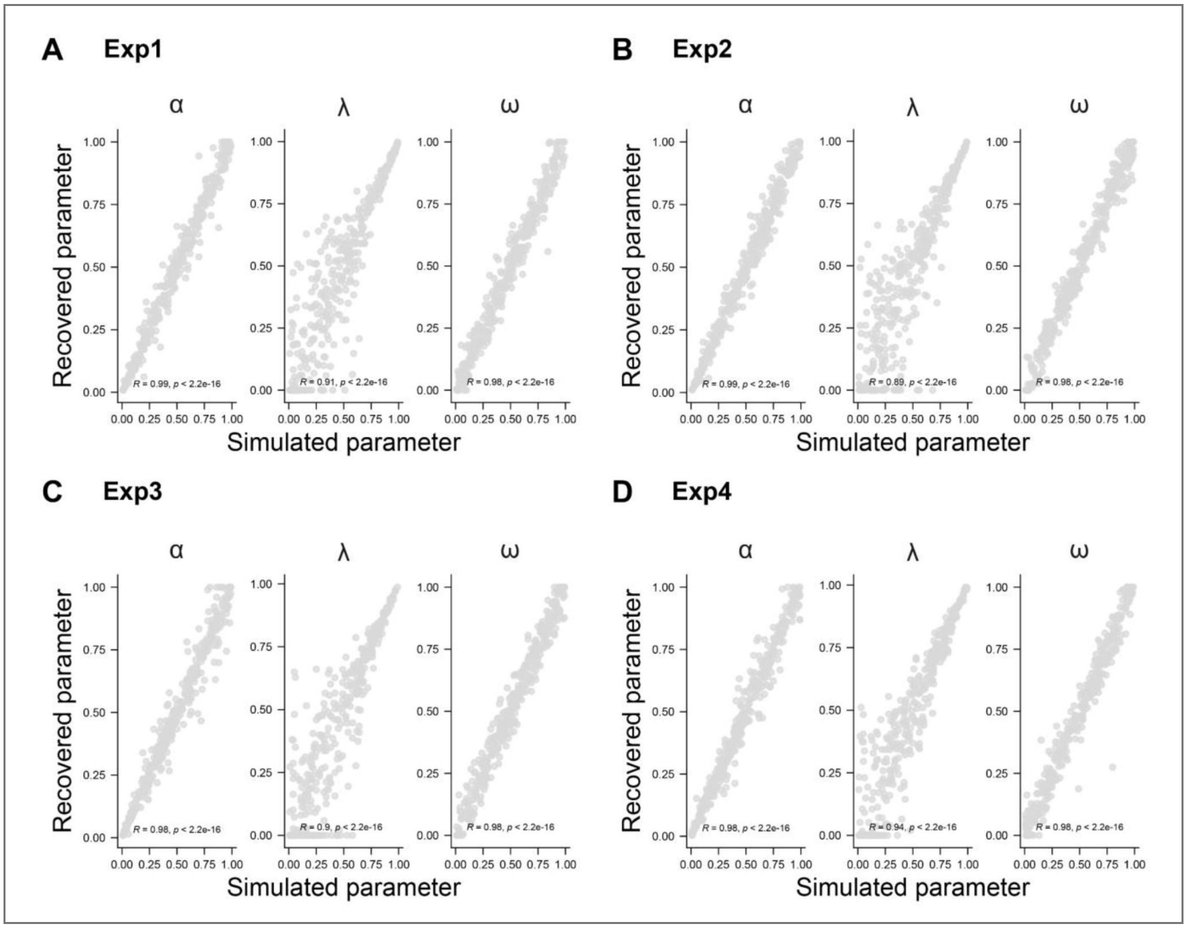



**Figure S5.** **Parameter recovery.**

To examine whether parameters of the winning model are recoverable, parameter recovery analyses were performed (**A - D**). For each experiment, we generated datasets from the winning model (N = 360 artificial agents; 85 simulated agents per condition) with added Gaussian noise. The winning model was then fitted to the simulated data. Parameter recovery was quantified by computing the correlation between true and fitted parameter values, with higher correlations indicating better recovery. Across all experiments, parameter recovery was strong: Spearman correlations between true and recovered values for the learning rate ***(α)***, relative weight on initial donations ***(λ)***, and social information weight (***w***) all exceeded 0.89.

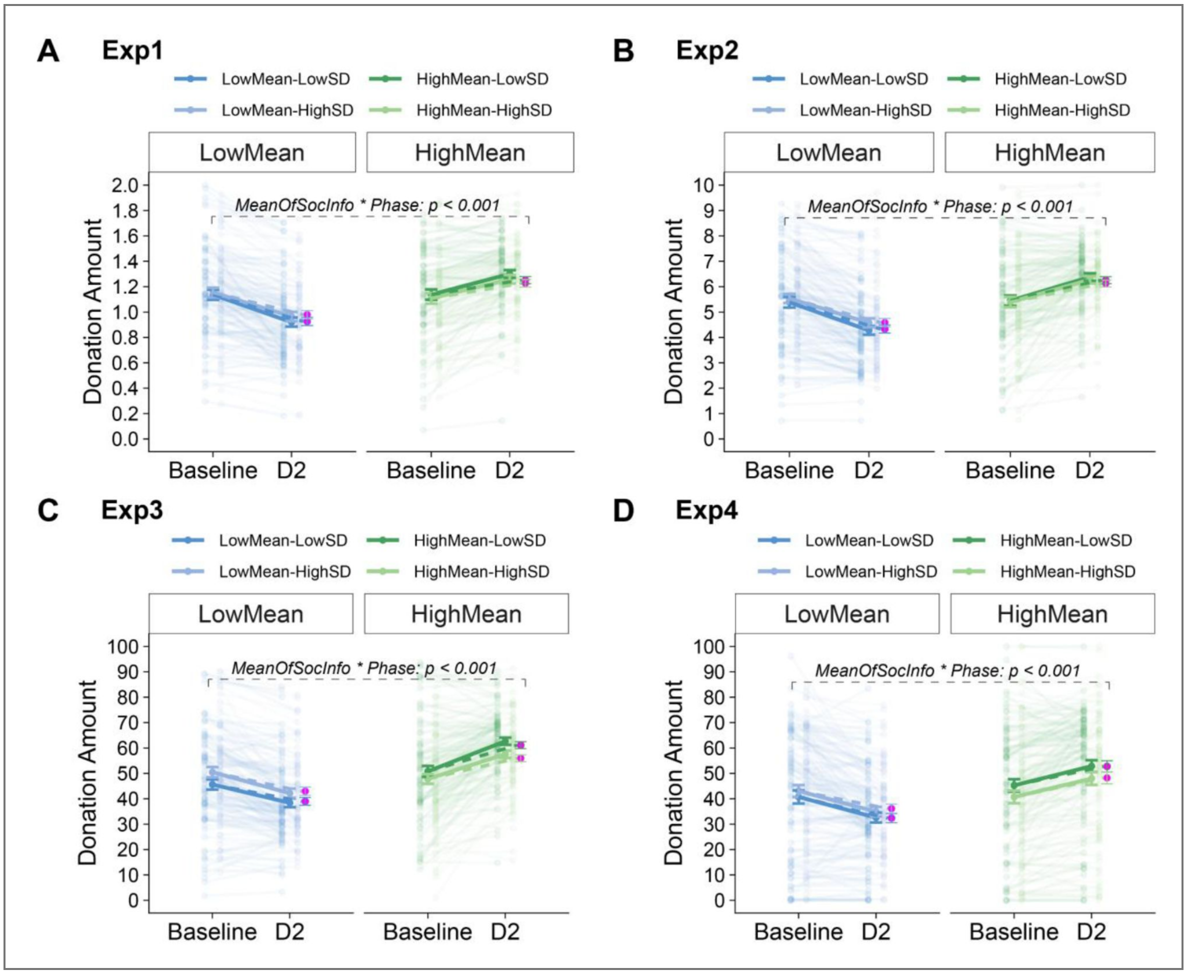



**Figure S6.** **Posterior predictive check: the predicted donation amount.**

The linear mixed-effects model conducted on donation amounts simulated from the winning model revealed a significant interaction between the mean of others' donations and the phase across all four experiments (Exp 1: $b$ = - 0.29, $SE$ = 0.021, $t(12613) = -13.96$, $p < 0.001$, 95% CI = [-0.34, -0.25]; Exp 2: $b$ = -1.68, $SE$ = 0.09, $t(13432) = -18.06$, $p < 0.001$, 95% CI = [-1.87, -1.50]; Exp 3: $b$ = -15.66, $SE$ = 0.94, $t(13627) = -16.74$, $p < 0.001$, 95% CI = [-17.49, -13.82]; Exp 4: $b$ = -14.44, $SE$ = 0.79, $t(14563) = -18.27$, $p < 0.001$, 95% CI = [-16.00, -12.90]). This pattern mirrors the real data, indicating that the winning model successfully captures the changes in the central tendency of individual donations. Solid lines indicate the observed donation behaviors, and dashed lines indicate model predictions generated using the best-fitting parameters of the winning model. Magenta dots represent the simulated values from the winning model.

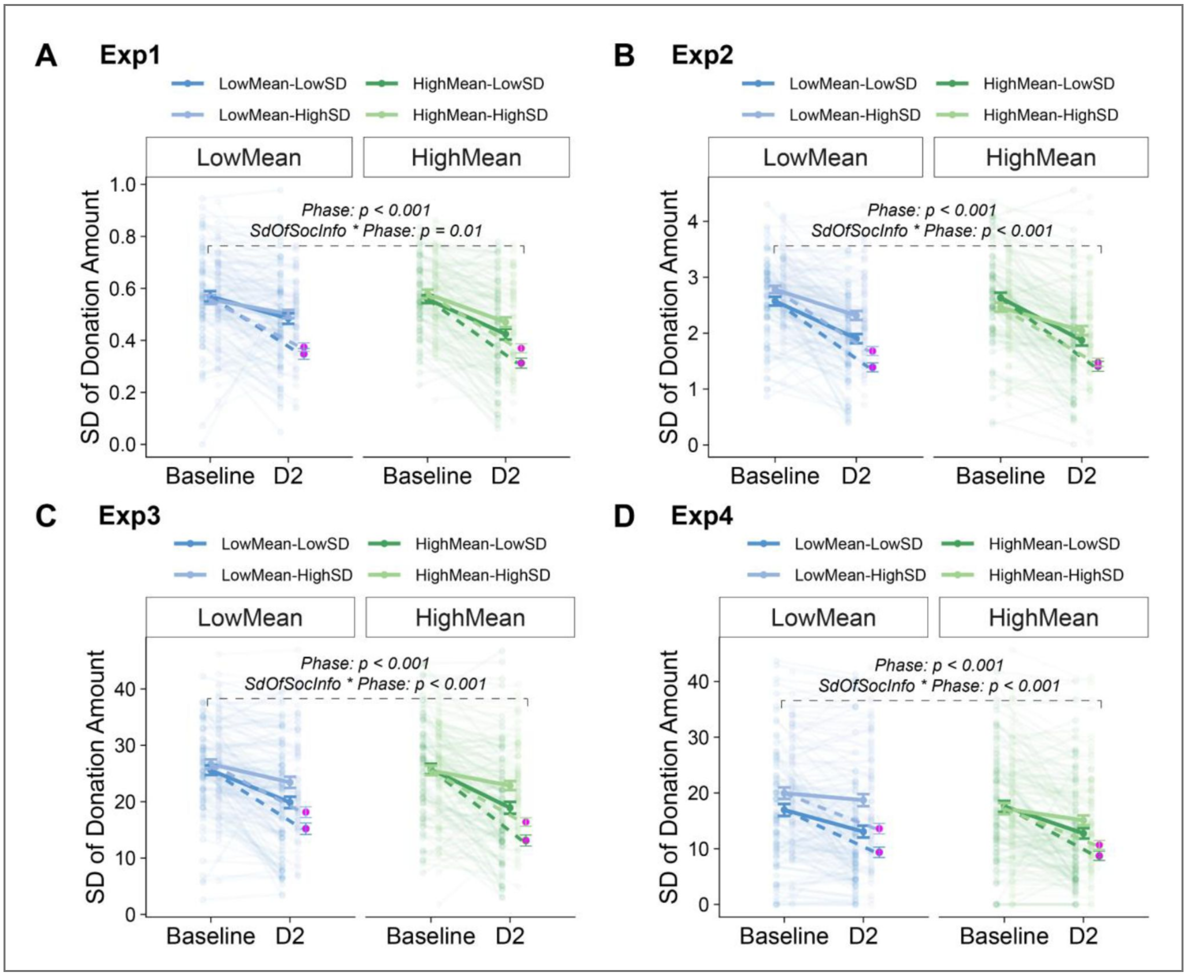



**Figure S7.** **Posterior predictive check: the standard deviation of the predicted donation amount.**

Mixed-effects ANOVAs conducted on the SD of predicted donation amounts revealed a significant main effect of the phase across all four experiments (Expl: $F(1, 320) = 833.875, p < 0.001$, partial $\eta^2 = 0.72$; Exp2: $F(1, 341) = 927.77, p < 0.001$, partial $\eta^2 = 0.73$; Exp3: $F(1, 346) = 766.36, p < 0.001$, partial $\eta^2 = 0.69$; Exp4: $F(1, 370) = 443.04, p < 0.001$, partial $\eta^2 = 0.55$). Moreover, the between the phase and the variability of others' donations was significant across all experiments (Exp1: $F(1, 320) = 6.34$, $p = 0.01$, partial $\eta^2 = 0.02$; Exp2: $F(1, 341) = 14.64, p < 0.001$, partial $\eta^2 = 0.04$; Exp3: $F(1, 346) = 18.44, p < 0.001$, partial $\eta^2 = 0.05$; Exp4: $F(1, 370) = 14.55, p < 0.001$, partial $\eta^2 = 0.04$). These results suggest that the winning model captures the variability dynamics observed in the real data. Solid lines indicate the observed donation behaviors, and dashed lines indicate model predictions generated using the best-fitting parameters of the winning model. Magenta dots represent the simulated values from the winning model.

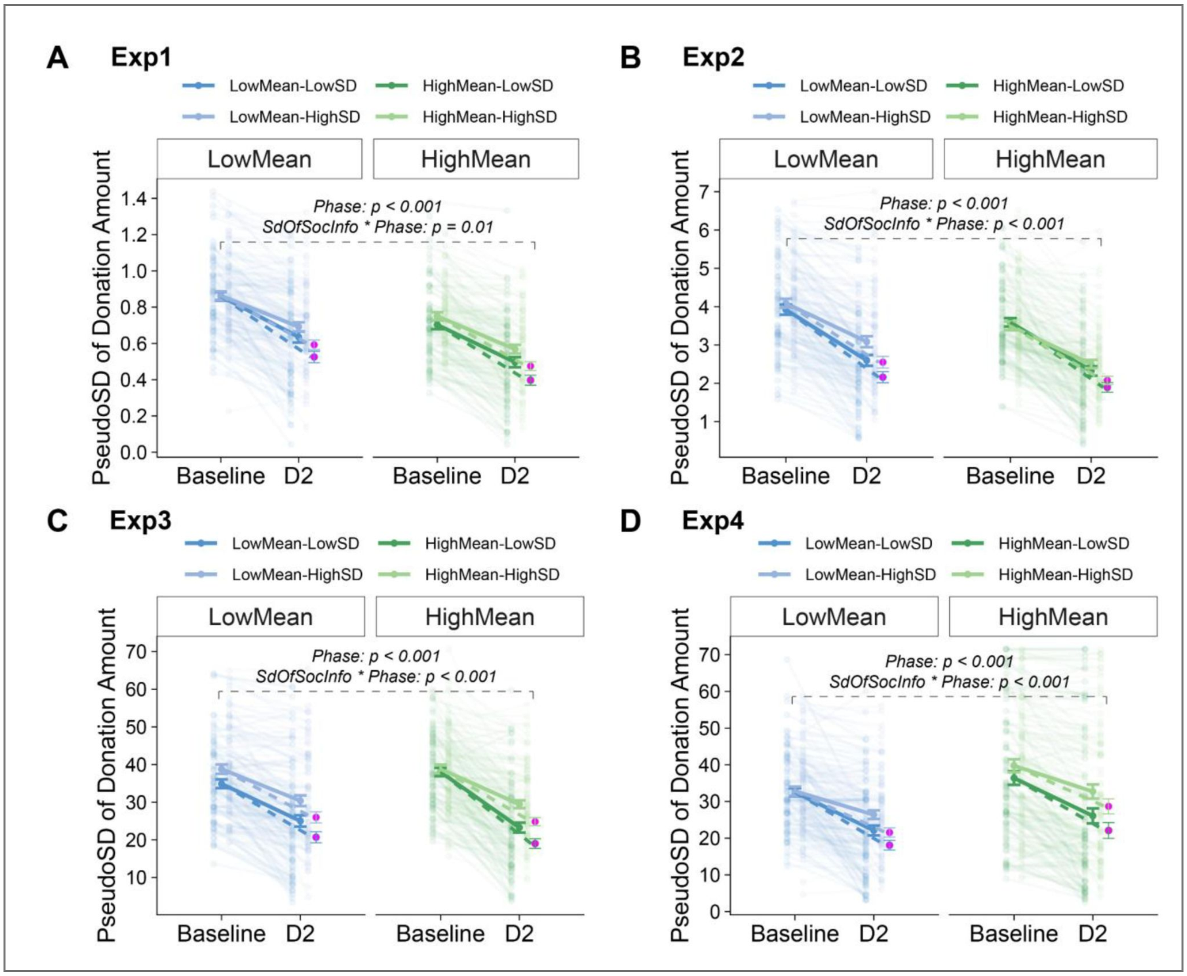



**Figure S8. Posterior predictive check: the Pseudo SD of the predicted donation amount.**

Mixedeffects ANOVAs conducted on the Pseudo SD of predicted donation amounts revealed a significant main effect of the phase across all four experiments (Exp1: $F(1, 320) = 833.875, p < 0.001$, partial $\eta^2 = 0.72$; Exp2: $F(1, 341) = 927.77, p < 0.001$, partial $\eta^2 = 0.73$; Exp3: $F(1, 346) = 766.36, p < 0.001$, partial $\eta^2 = 0.69$; Exp4: $F(1, 370) = 443.04, p < 0.001$, partial $\eta^2 = 0.55$). Moreover, we observed a significant interaction between the phase and the variability of others' donations (Exp1: $F(1, 320) = 6.34$, $p = 0.01$, partial $\eta^2 = 0.02$; Exp2: $F(1, 341) = 14.64, p < 0.001$, partial $\eta^2 = 0.04$; Exp3: $F(1, 346) = 18.44, p < 0.001$, partial $\eta^2 = 0.05$; Exp4: $F(1, 370) = 14.55, p < 0.001$, partial $\eta^2 = 0.04$). These results demonstrate that the winning model generates variability patterns closely matching those observed in the real data. Solid lines indicate the observed donation behaviors, and dashed lines indicate model predictions generated using the best-fitting parameters of the winning model. Magenta dots represent the simulated values from the winning model.

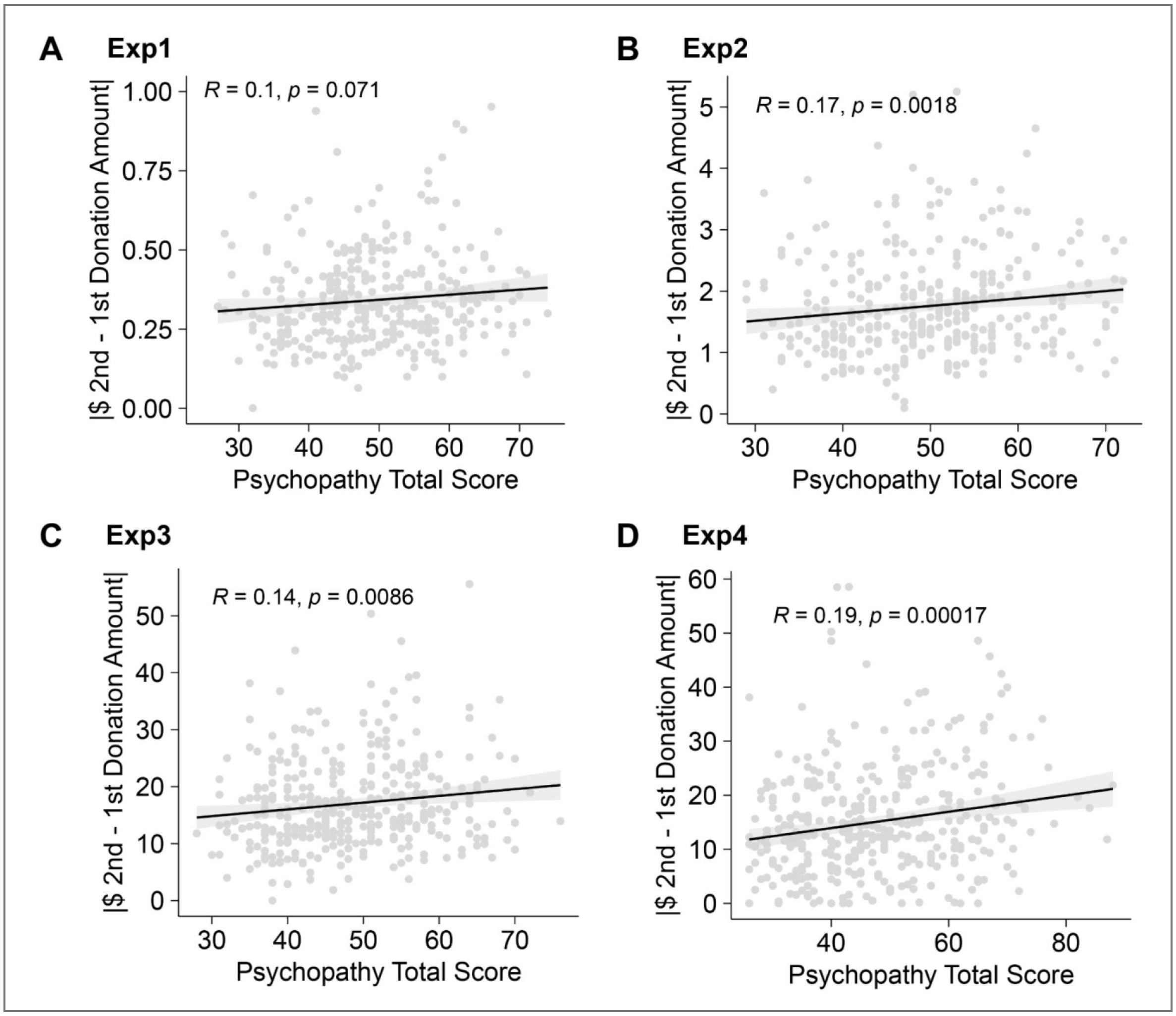



**Figure S9. Relationships between psychopathy and absolute donation changes from the first to the second individual donations, per experiment.**

The total psychopathy scores were positively correlated with the absolute donation changes in Experiment 2 - Experiment 4, with a marginally significant association observed in Experiment 1 (Expl: Spearman $\rho = 0.10$, $p = 0.071$, 95% CI: [-0.01, 0.21]; Exp2: Spearman$\rho = 0.17$,$p = 0.002$, 95% CI: [0.06, 0.27]; Exp3: Spearman $\rho = 0.14$,$p = 0.009$, 95% CI: [0.03, 0.24]; Exp4: $p = 0.19$,$p < 0.001$, 95% CI: [0.09, 0.29]).

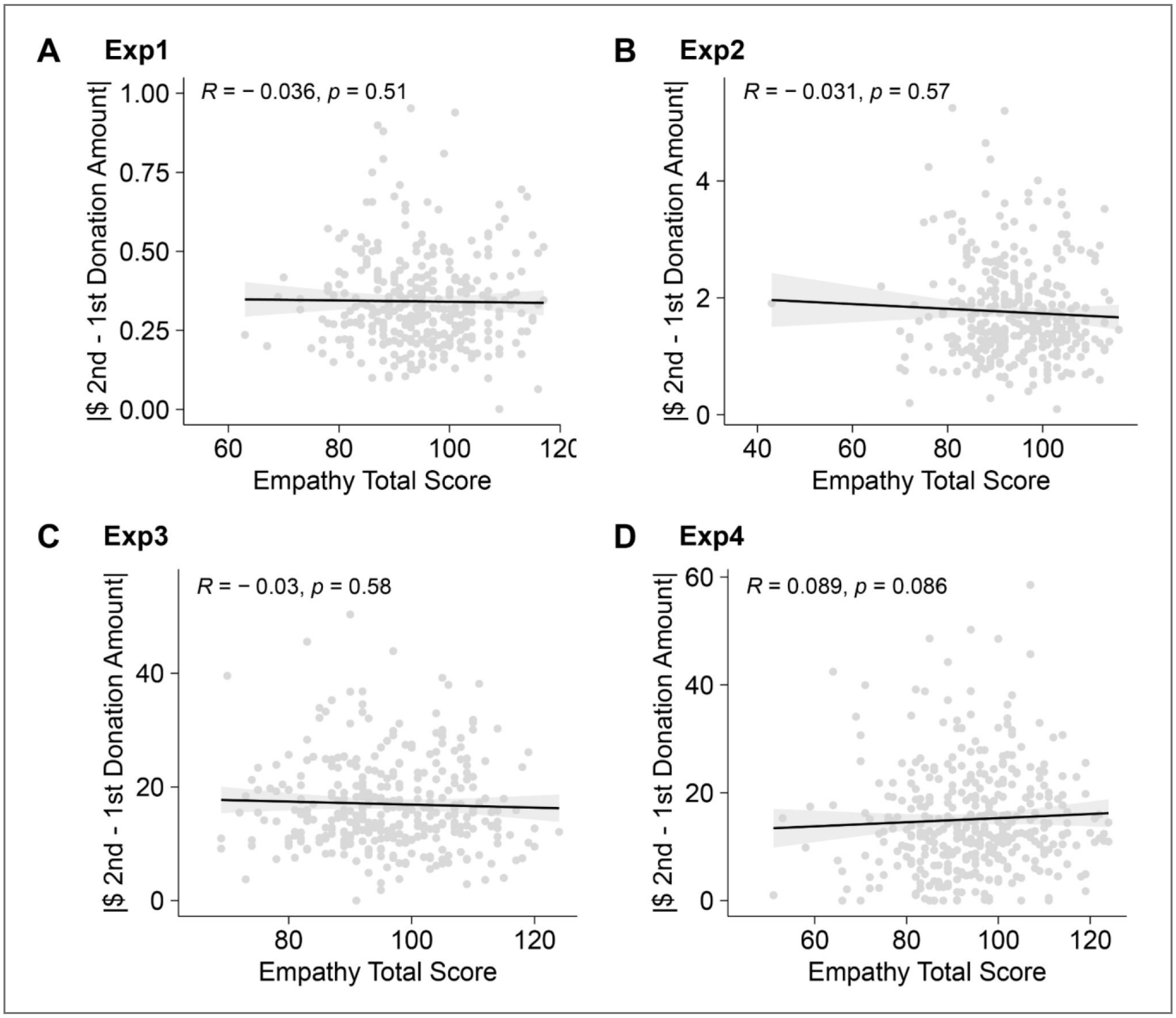



**Figure S10. Relationships between empathy and absolute donation changes from the first to the second individual donations.**

The total empathy scores showed no significant associations with the absolute donation changes across all experiments (Expl: Spearman $\rho$ = -0.04, $p$ = 0.51, 95% CI: [-0.15, 0.08]; Exp2: Spearman $\rho$ = -0.03, $p$ = 0.57, 95% CI: [-0.14, 0.08]; Exp3: Spearman ρ = -0.03, $p$ = 0.58, 95% CI: [-0.14, 0.08]; Exp4: $\rho$ = 0.09, $p$ = 0.086, 95% CI: [-0.02, 0.19]).

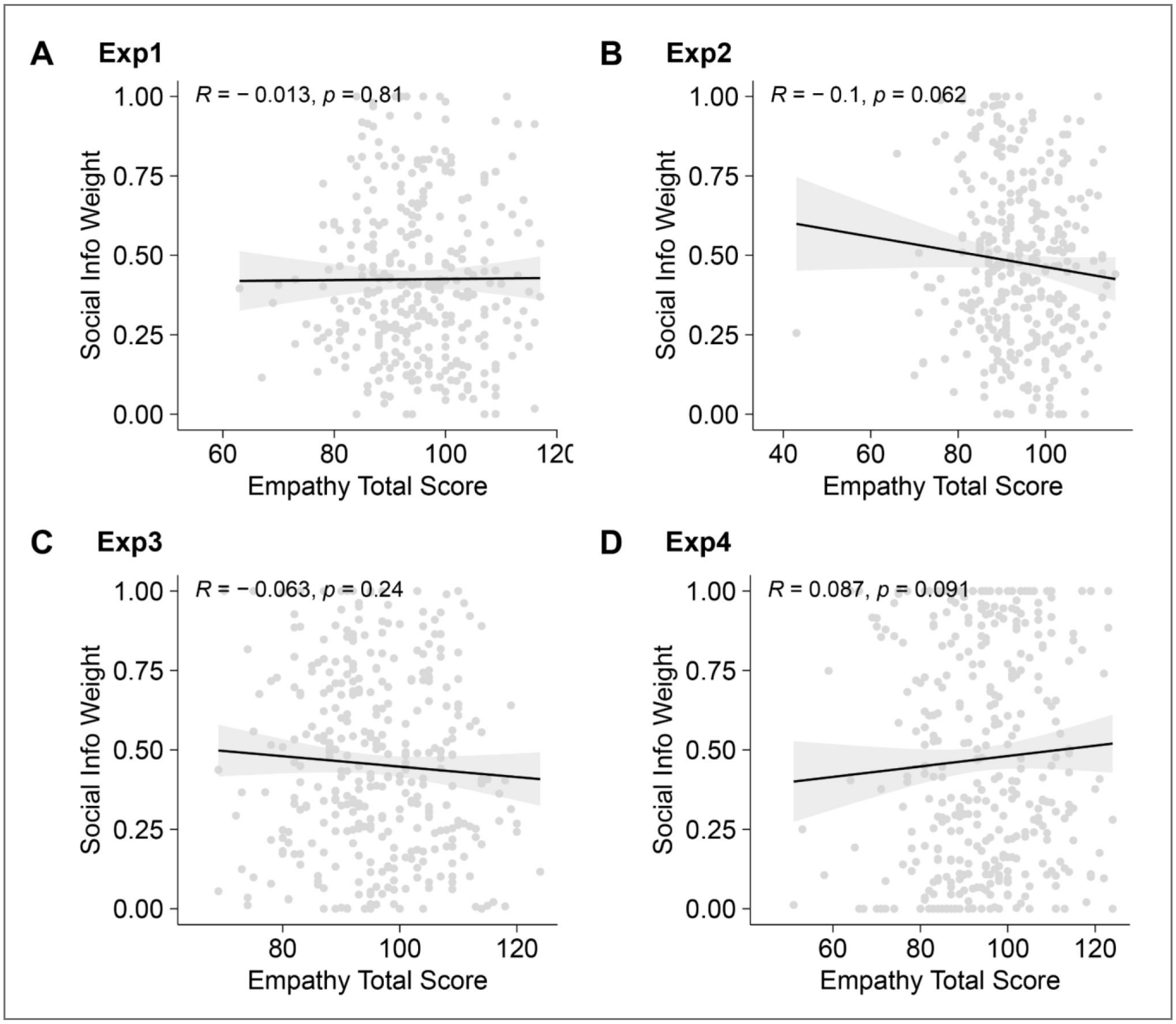



**Figure S11.** **Association between empathy and the social information weight parameters, per experiment.**

The total empathy scores were not significantly associated with the social information weight parameter (Expl: Spearman $\rho$ = -0.01, $p$ = 0.81, 95% CI: [-0.13, 0.10]; Exp2: Spearman $\rho$ = -0.10, $p$ = 0.062, 95% CI: [-0.21, 0.01]; Exp3: Spearman $\rho$ = -0.06, $p$ = 0.24, 95% CI: [-0.17, 0.05]; Exp4: $\rho$ = 0.09, $p$ = 0.091, 95% CI: [-0.02, 0.19]).

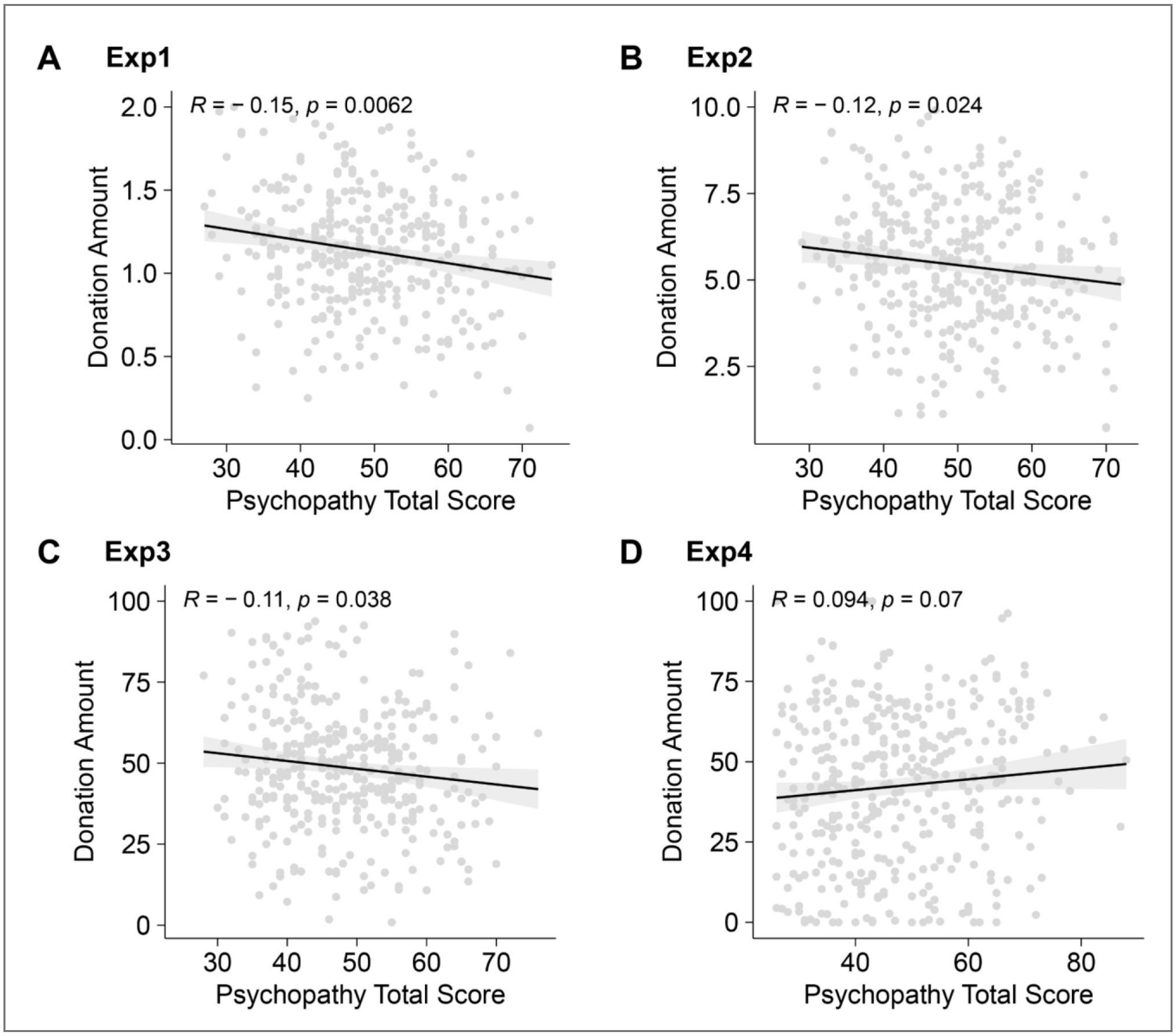



**Figure S12.** **Association between psychopathy and baseline donation amounts, per experiment.**

The total psychopathy scores were significantly negatively associated with the baseline donation amount in Experiment 1-3, but not in Experiment 4 (Exp1: Spearman $\rho$ = -0.15, $p$ = 0.006, 95% CI: [-0.26, -0.04]; Exp2: Spearman $\rho$ = -0.12, $p$ = 0.024, 95% CI: [-0.23, -0.01]; Exp3: Spearman ρ = -0.11, $p$ = 0.038, 95% CI: [-0.22, 0]; Exp4: $\rho$ = 0.09, $p$ = 0.07, 95% CI: [-0.01, 0.20]).

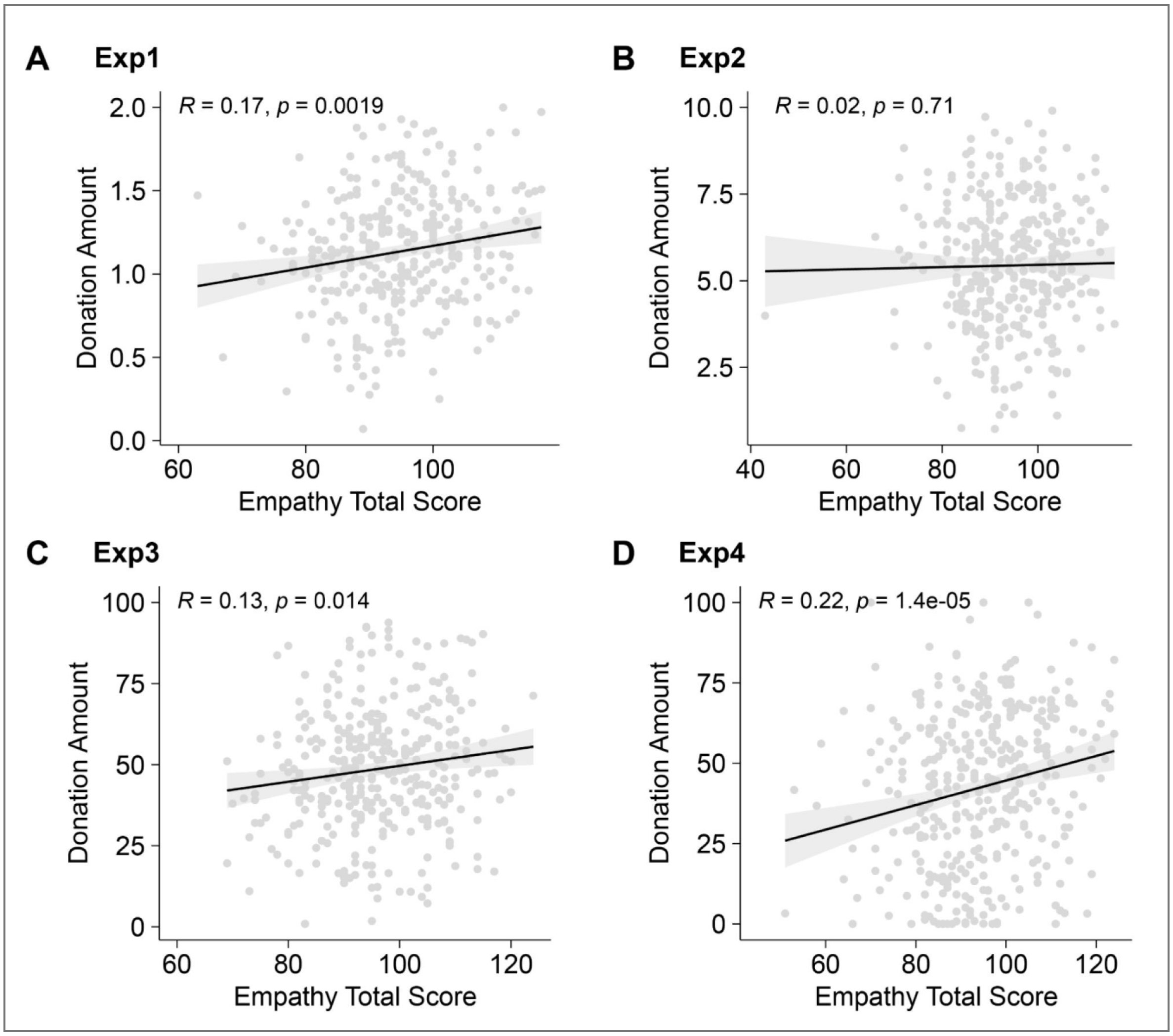



**Figure S13. Association between empathy and the baseline donation amounts, per experiment.**

Significant positive associations between the total empathy scores and the baseline donation amounts emerged in Experiment 1, 3 and 4, whereas this association was not observed in Experiment 2 (Expl: Spearman $\rho = 0.17$, $p = 0.002$, 95% CI: [0.06, 0.28]; Exp2: Spearman $\rho = 0.02$, $p = 0.71$, 95% CI: [-0.09, 0.13]; Exp3: Spearman $\rho = 0.13$,$p = 0.014$, 95% CI: [0.02, 0.24]; Exp4: $\rho = 0.22$,$p < 0.001$, 95% CI: [0.12, 0.32]).

**Figure S14.** **No association between psychopathy and the initial estimate of the number of animals.**

The total psychopathy scores were not associated with the first estimate of the number of animals in the perceptual-based social influence task (Spearman $\rho$ = -0.03, $p$ = 0.51, 95% CI: [-0.14, 0.07]).

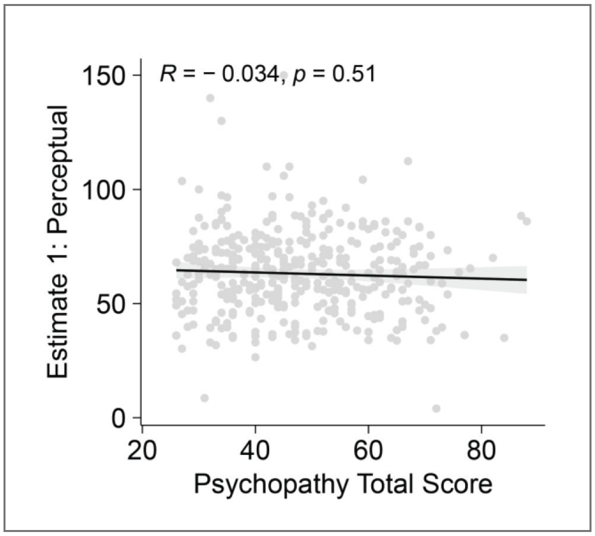
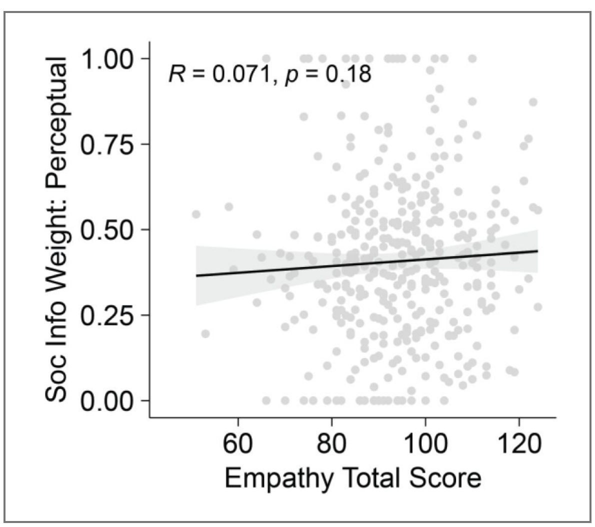



**Figure S15.** **No association between empathy and the social information weight in a perceptual task.**

The total empathy scores were not correlated with the social information weight in the perceptual-based social influence task (Spearman $\rho$ = 0.07, $p$ = 0.18, 95% CI: [-0.03, 0.17]).



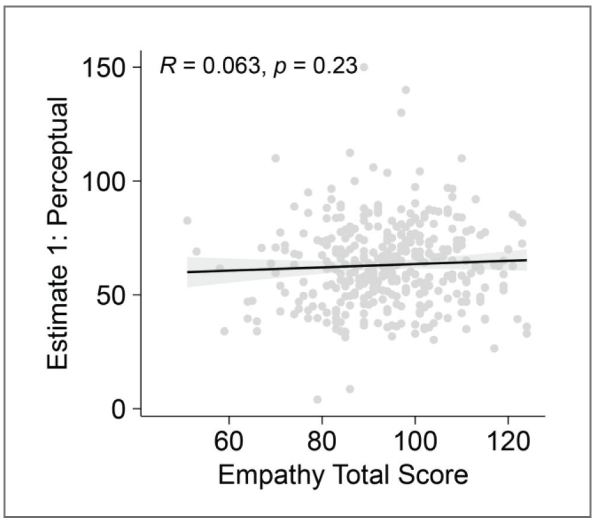



**Figure S16.** **No association between empathy and the initial estimate of the number of animals.**

The total empathy scores were not correlated with the first estimate of the number of animals in the perceptual-based social influence task (Spearman $\rho$ = 0.06, $p$ = 0.23, 95% CI: [-0.04, 0.17]).

## Supplementary Tables

| *Exp1: Fixed effect* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 1.11 | 0.08 | 32 | 14.38 | [0.96, 1.26] | < 0.001 |
| **Mean Cond** | 0.05 | 0.05 | 355 | 0.91 | [-0.05, 0.15] | 0.37 |
| **SD Cond** | 0.04 | 0.05 | 355 | 0.69 | [-0.07, 0.14] | 0.49 |
| **Phase** | 0.15 | 0.02 | 12652 | 9.08 | [0.12, 0.18] | < 0.001 |
| **Mean Cond * SD Cond** | -0.05 | 0.07 | 355 | -0.74 | [-0.20, 0.09] | 0.46 |
| **Mean Cond * Phase** | -0.34 | 0.02 | 12652 | -14.83 | [-0.39, -0.30] | < 0.001 |
| **SD Cond * Phase** | 0.01 | 0.02 | 12652 | 0.34 | [-0.04, 0.05] | 0.74 |
| **Mean Cond * SD Cond * Phase** | -0.04 | 0.03 | 12652 | -1.07 | [-0.10, 0.03] | 0.29 |

| *Exp2: Fixed effect* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 5.4 | 0.35 | 33 | 15.43 | [4.71, 6.07] | < 0.001 |
| **Mean Cond** | 0.14 | 0.24 | 372 | 0.59 | [-0.34, 0.63] | 0.55 |
| **SD Cond** | 0.06 | 0.24 | 372 | 0.24 | [-0.42, 0.54] | 0.81 |
| **Phase** | 0.88 | 0.07 | 13432 | 12.32 | [0.74, 1.02] | < 0.001 |
| **Mean Cond * SD Cond** | -0.21 | 0.35 | 372 | -0.60 | [-0.89, 0.47] | 0.55 |
| **Mean Cond * Phase** | -1.84 | 0.10 | 13432 | -18.16 | [-2.04, -1.64] | < 0.001 |
| **SD Cond * Phase** | 0.05 | 0.10 | 13432 | 0.50 | [-0.15, 0.25] | 0.62 |
| **Mean Cond * SD Cond * Phase** | -0.20 | 0.14 | 13432 | -1.38 | [-0.48, 0.08] | 0.17 |

| *Exp3: Fixed effect* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 47.81 | 0.35 | 32 | 13.27 | [40.74, 54.87] | < 0.001 |
| **Mean Cond** | 2.66 | 0.24 | 377 | 1.07 | [-2.21, 7.54] | 0.28 |
| **SD Cond** | 2.99 | 0.24 | 377 | 1.20 | [-1.88, 7.86] | 0.23 |
| **Phase** | 9.76 | 0.07 | 13627 | 13.75 | [8.37, 11.15] | < 0.001 |
| **Mean Cond * SD Cond** | -7.84 | 0.35 | 377 | -2.22 | [-14.76, -0.91] | 0.027 |
| **Mean Cond * Phase** | -18.03 | 0.10 | 13627 | -17.71 | [-20.03, -16.04] | < 0.001 |
| **SD Cond * Phase** | 2.09 | 0.10 | 13627 | 2.05 | [0.09, 4.09] | 0.040 |
| **Mean Cond * SD Cond * Phase** | -1.04 | 0.14 | 13627 | -0.72 | [-3.88, 1.79] | 0.47 |

| *Exp4: Fixed effect* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 40.71 | 2.69 | 154 | 15.16 | [35.44, 45.97] | < 0.001 |
| **Mean Cond** | 2.36 | 3.20 | 384 | 0.74 | [-3.01, 8.63] | 0.46 |
| **SD Cond** | 4.56 | 3.15 | 384 | 1.45 | [-1.62, 10.74] | 0.15 |
| **Phase** | 7.25 | 0.61 | 14563 | 11.92 | [6.06, 8.45] | < 0.001 |
| **Mean Cond * SD Cond** | -6.91 | 4.51 | 384 | -1.53 | [-15.75, 1.93] | 0.13 |
| **Mean Cond * Phase** | -15.42 | 0.87 | 14563 | -17.63 | [-17.13, -13.70] | < 0.001 |
| **SD Cond * Phase** | 0.33 | 0.86 | 14563 | 0.38 | [-1.36, 2.02] | 0.70 |
| **Mean Cond * SD Cond * Phase** | -0.30 | 1.23 | 14563 | -0.25 | [-2.72, 2.11] | 0.81 |

**Table S1. Linear mixed-effects model results predicting individual donation amounts as a function of the Mean and SD of Others' donations and Phase** *Individual donation amount ~ 1 + Mean Cond + SD Cond + Phase + Mean Cond * SD Cond + Mean Cond * Phase + SD Cond * Phase + Mean Cond * SD Cond * Phase + (1 | SubID) + (1 | Charity Item)*

| *Exp 1* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 321 | 20.18 | < 0.001 | 0.06 |
| **SD Cond** | 1 | 321 | 0.02 | 0.90 | < 0.001 |
| **Phase** | 1 | 321 | 4.57 | 0.03 | 0.01 |
| **Mean Cond * SD Cond** | 1 | 321 | 1.00 | 0.32 | 0.003 |

**Table S2.** **Mixed ANOVA on the averaged individual donation amount**

| | | | | | |
|---|---|---|---|---|---|
| **Mean Cond * Phase** | 1 | 321 | 255.26 | < 0.001 | 0.44 |
| **SD Cond * Phase** | 1 | 321 | 0.18 | 0.67 | 0.001 |
| **Mean Cond * SD Cond * Phase** | 1 | 321 | 0.60 | 0.44 | 0.002 |

Note: $g\ \eta^2$ - generalized eta squared

| ***Exp 2*** | ***df1*** | ***df2*** | ***F*** | ***p*** | ***partial $\eta^2$*** |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 341 | 29.95 | < 0.001 | 0.08 |
| **SD Cond** | 1 | 341 | 0.17 | 0.67 | 0.001 |
| **Phase** | 1 | 341 | 1.13 | 0.29 | 0.003 |
| **Mean Cond * SD Cond** | 1 | 341 | 0.82 | 0.37 | 0.002 |
| **Mean Cond * Phase** | 1 | 341 | 258.33 | < 0.001 | 0.43 |
| **SD Cond * Phase** | 1 | 341 | 0.16 | 0.69 | < 0.001 |
| **Mean Cond * SD Cond * Phase** | 1 | 341 | 0.68 | 0.41 | 0.002 |

| ***Exp 3*** | ***df1*** | ***df2*** | ***F*** | ***p*** | ***partial $\eta^2$*** |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 346 | 37.07 | < 0.001 | 0.10 |
| **SD Cond** | 1 | 346 | 0.01 | 0.93 | < 0.001 |
| **Phase** | 1 | 346 | 6.25 | 0.01 | 0.02 |
| **Mean Cond * SD Cond** | 1 | 346 | 5.84 | 0.02 | 0.02 |
| **Mean Cond * Phase** | 1 | 346 | 231.13 | < 0.001 | 0.40 |
| **SD Cond * Phase** | 1 | 346 | 1.65 | 0.20 | 0.005 |
| **Mean Cond * SD Cond * Phase** | 1 | 346 | 0.18 | 0.70 | 0.001 |

| ***Exp 4*** | ***df1*** | ***df2*** | ***F*** | ***p*** | ***partial $\eta^2$*** |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 370 | 15.81 | < 0.001 | 0.04 |
| **SD Cond** | 1 | 370 | 0.28 | 0.59 | 0.001 |

**Table S2.** (continued)

| | | | | | |
|---|---|---|---|---|---|
| **Phase** | 1 | 370 | 0.31 | 0.58 | 0.001 |
| **Mean Cond * SD Cond** | 1 | 370 | 2.50 | 0.12 | 0.007 |
| **Mean Cond * Phase** | 1 | 370 | 138.10 | < 0.001 | 0.27 |
| **SD Cond * Phase** | 1 | 370 | 0.02 | 0.90 | < 0.001 |
| **Mean Cond * SD Cond * Phase** | 1 | 370 | 0.01 | 0.91 | < 0.001 |

**Table S2.** (continued)

**Table S3.** **Control analysis: The linear mixed-effects model predicting individual donation amounts as a function of the Mean and SD of Others' donations in the Baseline Phase**

*Individual donation amount ~ 1 + Mean Cond + SD Cond + Mean Cond * SD Cond + (1 | SubiD) + (1 | Charity Item)*

| ***Exp1: Fixed effect*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 1.11 | 0.09 | 31 | 13.00 | [0.94, 1.27] | < 0.001 |
| **Mean Cond** | 0.05 | 0.06 | 321 | 0.82 | [-0.06, 0.16] | 0.41 |
| **SD Cond** | 0.04 | 0.06 | 321 | 0.63 | [-0.08, 0.15] | 0.53 |
| **Mean Cond * SD Cond** | -0.05 | 0.08 | 321 | -0.67 | [-0.21, 0.10] | 0.50 |

| ***Exp2: Fixed effect*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 5.39 | 0.40 | 32 | 13.39 | [4.60, 6.18] | < 0.001 |
| **Mean Cond** | 0.14 | 0.28 | 341 | 0.52 | [-0.40, 0.69] | 0.60 |
| **SD Cond** | 0.06 | 0.28 | 341 | 0.21 | [-0.49, 0.61] | 0.84 |
| **Mean Cond * SD Cond** | -0.21 | 0.40 | 341 | -0.53 | [-0.99, 0.57] | 0.60 |

| ***Exp3: Fixed effect*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 47.81 | 4.07 | 32 | 11.73 | [39.82, 55.80] | < 0.001 |
| **Mean Cond** | 2.66 | 2.86 | 346 | 0.93 | [-2.94, 8.26] | 0.35 |
| **SD Cond** | 2.99 | 2.86 | 346 | 1.05 | [-2.61, 8.59] | 0.29 |
| **Mean Cond * SD Cond** | -7.84 | 4.06 | 346 | -1.93 | [-15.80, 0.12] | 0.054 |

| ***Exp4: Fixed effect*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 40.71 | 3.03 | 126 | 13.44 | [34.77, 46.64] | < 0.001 |
| **Mean Cond** | 2.36 | 3.50 | 370 | 0.67 | [-4.50, 9.21] | 0.50 |
| **SD Cond** | 4.56 | 3.45 | 370 | 1.32 | [-2.20, 11.32] | 0.19 |
| **Mean Cond * SD Cond** | -6.91 | 4.93 | 370 | -1.40 | [-16.58, 2.76] | 0.16 |

| *Exp 1* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 321 | 1.40 | 0.24 | 0.004 |
| **SD Cond** | 1 | 321 | 1.07 | 0.30 | 0.003 |
| **Phase** | 1 | 321 | 146.63 | < 0.001 | 0.31 |
| **Mean Cond * SD Cond** | 1 | 321 | 0.76 | 0.38 | 0.002 |
| **Mean Cond * Phase** | 1 | 321 | 10.17 | 0.002 | 0.03 |
| **SD Cond * Phase** | 1 | 321 | 2.87 | 0.09 | 0.009 |
| **Mean Cond * SD Cond * Phase** | 1 | 321 | 0.02 | 0.89 | < 0.001 |

Note: *partial η²* - partial eta squared

| *Exp 2* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 341 | 3.62 | 0.06 | 0.01 |
| **SD Cond** | 1 | 341 | 4.51 | 0.03 | 0.01 |
| **Phase** | 1 | 341 | 260.99 | < 0.001 | 0.43 |
| **Mean Cond * SD Cond** | 1 | 341 | 4.44 | 0.04 | 0.01 |
| **Mean Cond * Phase** | 1 | 341 | 0.11 | 0.74 | < 0.001 |
| **SD Cond * Phase** | 1 | 341 | 14.64 | < 0.001 | 0.04 |
| **Mean Cond * SD Cond * Phase** | 1 | 341 | 0.88 | 0.35 | 0.003 |

| *Exp 3* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 346 | 0.55 | 0.46 | 0.002 |

**Table S4.** **Mixed ANOVA on the SD of the individual donation amounts**

| | | | | | |
|---|---|---|---|---|---|
| **SD Cond** | 1 | 346 | 6.32 | 0.01 | 0.02 |
| **Phase** | 1 | 346 | 144.57 | < 0.001 | 0.30 |
| **Mean Cond * SD Cond** | 1 | 341 | 0.08 | 0.77 | < 0.001 |
| **Mean Cond * Phase** | 1 | 346 | 0.16 | 0.69 | < 0.001 |
| **SD Cond * Phase** | 1 | 346 | 18.44 | < 0.001 | 0.05 |
| **Mean Cond * SD Cond * Phase** | 1 | 346 | 1.20 | 0.28 | 0.003 |

| ***Exp 4*** | ***df1*** | ***df2*** | ***F*** | ***p*** | ***partial η²*** |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 370 | 2.63 | 0.11 | 0.007 |
| **SD Cond** | 1 | 370 | 7.99 | 0.005 | 0.02 |
| **Phase** | 1 | 370 | 73.18 | < 0.001 | 0.17 |
| **Mean Cond * SD Cond** | 1 | 370 | 3.20 | 0.07 | 0.009 |
| **Mean Cond * Phase** | 1 | 370 | 1.58 | 0.21 | 0.004 |
| **SD Cond * Phase** | 1 | 370 | 14.55 | < 0.001 | 0.04 |
| **Mean Cond * SD Cond * Phase** | 1 | 370 | 0.005 | 0.95 | < 0.001 |

**Table S4.** (continued)

| *Exp 1* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 321 | 33.46 | < 0.001 | 0.09 |
| **SD Cond** | 1 | 321 | 3.76 | 0.054 | 0.01 |
| **Phase** | 1 | 321 | 372.79 | < 0.001 | 0.54 |
| **Mean Cond * SD Cond** | 1 | 321 | 0.39 | 0.53 | 0.001 |
| **Mean Cond * Phase** | 1 | 321 | 0.05 | 0.83 | < 0.001 |
| **SD Cond * Phase** | 1 | 321 | 3.56 | 0.06 | 0.01 |
| **Mean Cond * SD Cond * Phase** | 1 | 321 | 0.52 | 0.47 | 0.002 |

Note: *partial* $\eta^2$ - partial eta squared

**Table S5.** **Mixed ANOVA on the '*pseudo-SD of individual donation' amounts***

| *Exp 2* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 341 | 15.77 | < 0.001 | 0.04 |
| **SD Cond** | 1 | 341 | 2.75 | 0.10 | 0.008 |
| **Phase** | 1 | 341 | 464.22 | < 0.001 | 0.58 |
| **Mean Cond * SD Cond** | 1 | 341 | 1.62 | 0.20 | 0.005 |
| **Mean Cond * Phase** | 1 | 341 | 0.08 | 0.77 | < 0.001 |
| **SD Cond * Phase** | 1 | 341 | 7.82 | 0.005 | 0.02 |
| **Mean Cond * SD Cond * Phase** | 1 | 341 | 0.04 | 0.84 | < 0.001 |

| *Exp 3* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 346 | 0.01 | 0.91 | < 0.001 |
| **SD Cond** | 1 | 346 | 13.06 | < 0.001 | 0.04 |
| **Phase** | 1 | 346 | 397.08 | < 0.001 | 0.53 |
| **Mean Cond * SD Cond** | 1 | 341 | 0.26 | 0.61 | 0.001 |
| **Mean Cond * Phase** | 1 | 346 | 7.56 | 0.006 | 0.02 |
| **SD Cond * Phase** | 1 | 346 | 10.43 | 0.001 | 0.03 |
| **Mean Cond * SD Cond * Phase** | 1 | 346 | 3.44 | 0.064 | 0.01 |

| *Exp 4* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 370 | 12.18 | 0.001 | 0.03 |
| **SD Cond** | 1 | 370 | 5.80 | 0.02 | 0.02 |
| **Phase** | 1 | 370 | 238.73 | < 0.001 | 0.39 |
| **Mean Cond * SD Cond** | 1 | 370 | 0.79 | 0.37 | 0.002 |
| **Mean Cond * Phase** | 1 | 370 | 0.06 | 0.81 | < 0.001 |
| **SD Cond * Phase** | 1 | 370 | 10.72 | 0.001 | 0.03 |

**Table S5.** (continued)

**Table S5.** (continued)

| | | | | | |
|---|---|---|---|---|---|
| **Mean Cond * SD Cond * Phase** | 1 | 370 | 0.11 | 0.74 | < 0.001 |

**Table S6. Linear mixed-effects model predicting individual novel donation amounts as a function of the Mean and SD of Others' donations**

*Individual novel donation amount ~ 1 + Mean Cond + SD Cond + Mean Cond * SD Cond + (1 | SubID) + (1 | Charity Item)*

| *Exp4: Fixed effect* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 33.15 | 4.32 | 8 | 7.67 | [24.69, 41.61] | < 0.001 |
| **Mean Cond** | 16.94 | 3.37 | 370 | 5.03 | [10.34, 23.54] | < 0.001 |
| **SD Cond** | -1.76 | 3.42 | 370 | -0.51 | [-8.45, 4.94] | 0.61 |
| **Mean Cond * SD Cond** | -4.22 | 4.78 | 370 | -0.88 | [-13.59, 5.15] | 0.38 |

**Table S7. Mixed ANOVA on the standard deviation *of individual novel donation amounts***

| *Exp 4* | *df1* | *df2* | *F* | *p* | *partial η²* |
|---|---|---|---|---|---|
| **Mean Cond** | 1 | 370 | 0.72 | 0.40 | 0.002 |
| **SD Cond** | 1 | 370 | 9.87 | 0.002 | 0.026 |
| **Mean Cond * SD Cond** | 1 | 370 | 2.20 | 0.14 | 0.006 |

Note: *partial η²* - partial eta squared

**Table S8.** **Summary table of model comparison results.**

AIC and BIC were calculated and then summed across participants. PXP (protected exceedance probability) was calculated among all models. Model 0 is the Intercept Only Model; Model 1 is the D1 Only Model; Model 2 is the Prediction Only Model; Model 2b is a variant of the Prediction Only model with a free initial prediction; Model 3 is a Hybrid of D1 and Prediction Model; Model 3b is a variant of Hybrid of D1 and Prediction Model with a free initial prediction. #param means the number of parameters.

| | | Model 0 | Model 1 | Model 2 | Model 2b | Model 3 | Model 3b |
|---|---|---|---|---|---|---|---|
| | #param | 2 | 4 | 4 | 4 | 3 | 3 |
| Exp 1 | AIC | -101626.2 | -117195.5 | -109075.8 | -104368.6 | **-118502.3** | -114206.5 |
| | BIC | -99658.2 | -113259.5 | -105139.8 | -100432.6 | **-115550.3** | -111254.5 |
| | PXP (%) | 0 | 0 | 0 | 0 | **100** | 0 |
| Exp 2 | AIC | 28341.9 | 14015.4 | 20794.7 | 25866.0 | **12311.3** | 17084.1 |
| | BIC | 30265.3 | 17862.1 | 24641.5 | 29712.8 | **15196.3** | 19969.1 |
| | PXP (%) | 0 | 0 | 0 | 0 | **100** | 0 |
| Exp 3 | AIC | 222483.0 | 206804.2 | 214030.8 | 219738.6 | **203516.5** | 209481.5 |
| | BIC | 224434.3 | 210706.8 | 217933.3 | 223641.1 | **206443.4** | 212408.4 |
| | PXP (%) | 0 | 0 | 0 | 0 | **100** | 0 |
| Exp 4 | AIC | 213502.8 | 204673.4 | 213189.6 | 209286.2 | 208159.1 | **203790.8** |
| | BIC | 215587.9 | 208843.5 | 217359.7 | 213456.4 | 211286.7 | **206918.3** |
| | PXP (%) | 0 | 0.03 | 0 | 0 | **67.32** | 32.65 |

| *Exp1* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.25 | 0.073 | 325 | 3.43 | [0.108, 0.396] | < 0.001 |
| **Psychopathy** | 0.002 | 0.001 | 318 | 2.00 | [0.000, 0.003] | 0.045 |
| **Age** | -0.001 | 0.002 | 318 | -0.44 | [-0.006, 0.004] | 0.66 |
| **Gender** | 0.012 | 0.019 | 318 | 0.66 | [-0.024, 0.049] | 0.51 |
| **Mean Cond** | 0.018 | 0.023 | 318 | 0.77 | [-0.028, 0.064] | 0.44 |
| **SD Cond** | 0.004 | 0.024 | 318 | 0.18 | [-0.042, 0.050] | 0.86 |
| **Mean Cond * SD Cond** | 0.026 | 0.033 | 318 | 0.80 | [-0.038, 0.091] | 0.42 |

Note: Gender was coded as non-Female (0) and Female (1).

| *Exp2* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 1.300 | 0.310 | 343 | 4.43 | [0.735, 1.949] | < 0.001 |
| **Psychopathy** | 0.011 | 0.005 | 338 | 2.28 | [0.001, 0.020] | 0.023 |
| **Age** | -0.007 | 0.005 | 338 | -1.46 | [-0.016, 0.002] | 0.14 |

**Table S9. Linear mixed-effects model predicting absolute donation changes from psychopathy scores, after controlling for confounding variables**

*|D2 — D1|~1 + Psychopathy Score + Age + Gender + Mean Cond + SD Cond + + Mean Cond * SD Cond + (1|SubID) + (1|Charity Item)*

| | | | | | | |
|---|---|---|---|---|---|---|
| **Gender** | -0.110 | 0.120 | 338 | -0.94 | [-0.346, 0.122] | 0.35 |
| **Mean Cond** | 0.140 | 0.120 | 338 | 1.11 | [-0.106, 0.382] | 0.37 |
| **SD Cond** | 0.110 | 0.120 | 338 | 0.87 | [-0.137, 0.352] | 0.39 |
| **Mean Cond * SD Cond** | -0.042 | 0.180 | 338 | -0.23 | [-0.388, 0.305] | 0.81 |

| ***Exp3*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 14.00 | 4.400 | 344 | 3.23 | [5.544, 22.633] | < 0.001 |
| **Psychopathy** | 0.140 | 0.045 | 343 | 3.06 | [0.049, 0.225] | 0.002 |
| **Age** | -0.280 | 0.150 | 343 | -1.81 | [-0.583, 0.022] | 0.070 |
| **Gender** | 2.400 | 1.100 | 343 | 2.2 | [0.269, 4.605] | 0.028 |
| **Mean Cond** | -1.300 | 1.200 | 343 | -1.12 | [-3.654, 0.993] | 0.26 |
| **SD Cond** | 2.700 | 1.200 | 343 | 2.29 | [0.391, 5.030] | 0.022 |
| **Mean Cond * SD Cond** | -2.900 | 1.700 | 343 | -1.73 | [-6.209, 0.385] | 0.083 |

| ***Exp4*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 8.900 | 3.000 | 367 | 2.92 | [2.936, 14.879] | 0.003 |
| **Psychopathy** | 0.140 | 0.040 | 367 | 3.56 | [0.064, 0.222] | < 0.001 |
| **Age** | -0.057 | 0.041 | 367 | -1.38 | [-0.139, 0.024] | 0.17 |
| **Gender** | 1.400 | 1.000 | 367 | 1.35 | [-0.644, 3.461] | 0.18 |
| **Mean Cond** | 0.780 | 1.500 | 367 | 0.54 | [-2.077, 3.640] | 0.59 |
| **SD Cond** | 1.200 | 1.400 | 367 | 0.84 | [-1.610, 4.025] | 0.40 |
| **Mean Cond * SD Cond** | -0.330 | 2.100 | 367 | -0.16 | [-4.381, 3.728] | 0.87 |

**Table S9.** (continued)

| *Exp1* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.439 | 0.135 | 315 | 3.24 | [0.173, 0.705] | 0.001 |
| **Psychopathy** | 0.002 | 0.001 | 315 | 1.70 | [0.000, 0.005] | 0.090 |
| $\alpha$ | 0.043 | 0.085 | 315 | 0.51 | [-0.124, 0.211] | 0.609 |
| $\lambda$ | -0.057 | 0.045 | 315 | -1.25 | [-0.146, 0.032] | 0.211 |
| **Age** | -0.006 | 0.004 | 315 | -1.57 | [-0.014, 0.002] | 0.117 |
| **Gender** | 0.026 | 0.032 | 315 | 0.81 | [-0.037, 0.088] | 0.417 |
| **Mean Cond** | -0.050 | 0.039 | 315 | -1.26 | [-0.127, 0.028] | 0.209 |
| **SD Cond** | 0.051 | 0.042 | 315 | 1.24 | [-0.030, 0.133] | 0.217 |
| **Mean Cond * SD Cond** | -0.013 | 0.056 | 315 | -0.23 | [-0.122, 0.096] | 0.815 |

Note: Gender was coded as non-Female (0) and Female (1).

| *Exp2* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.412 | 0.105 | 336 | 3.93 | [0.206, 0.618] | < 0.001 |
| **Psychopathy** | 0.004 | 0.001 | 336 | 3.04 | [0.002, 0.007] | 0.003 |
| $\alpha$ | -0.298 | 0.095 | 336 | -3.13 | [-0.485, -0.111] | 0.002 |
| $\lambda$ | -0.011 | 0.048 | 336 | -0.24 | [-0.105, 0.082] | 0.811 |
| **Age** | -0.003 | 0.001 | 336 | -1.78 | [-0.005, 0.000] | 0.076 |
| **Gender** | -0.010 | 0.037 | 336 | -0.26 | [-0.083, 0.064] | 0.793 |
| **Mean Cond** | -0.039 | 0.039 | 336 | -1.01 | [-0.116, 0.037] | 0.312 |
| **SD Cond** | 0.080 | 0.042 | 336 | 1.91 | [-0.002, 0.163] | 0.057 |
| **Mean Cond * SD Cond** | 0.018 | 0.055 | 336 | 0.32 | [-0.091, 0.127] | 0.748 |

| *Exp3* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.420 | 0.161 | 341 | 2.61 | [0.104, 0.735] | 0.009 |
| **Psychopathy** | 0.004 | 0.002 | 341 | 2.43 | [0.001, 0.007] | 0.015 |

**Table S10. Linear regression predicting the social information weight from psychopathy scores, after controlling for other modeling parameters and confounding variables**

*w ~ 1 + Psychopathy Score + α + λ + Age + Gender + Mean Cond + SD Cond + + Mean Cond * SD Cond*

| | | | | | | |
|---|---|---|---|---|---|---|
| $\alpha$ | -0.144 | 0.090 | 341 | -1.59 | [-0.321, 0.034] | 0.113 |
| $\lambda$ | -0.072 | 0.051 | 341 | -1.43 | [-0.172, 0.027] | 0.155 |
| **Age** | -0.007 | 0.005 | 341 | -1.28 | [-0.018, 0.004] | 0.201 |
| **Gender** | 0.067 | 0.039 | 341 | 1.72 | [-0.009, 0.143] | 0.086 |
| **Mean Cond** | -0.030 | 0.042 | 341 | -0.72 | [-0.112, 0.052] | 0.475 |
| **SD Cond** | 0.149 | 0.047 | 341 | 3.18 | [0.057, 0.242] | 0.002 |
| **Mean Cond * SD Cond** | -0.058 | 0.059 | 341 | -0.98 | [-0.174, 0.058] | 0.326 |

| ***Exp4*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 0.264 | 0.106 | 365 | 2.50 | [0.056, 0.472] | 0.013 |
| **Psychopathy** | 0.007 | 0.001 | 365 | 5.27 | [0.004, 0.010] | < 0.001 |
| $\alpha$ | -0.389 | 0.091 | 365 | -4.29 | [-0.567, -0.211] | < 0.001 |
| $\lambda$ | -0.023 | 0.049 | 365 | -0.48 | [-0.120, 0.073] | 0.635 |
| **Age** | -0.002 | 0.001 | 365 | -1.64 | [-0.005, 0.000] | 0.101 |
| **Gender** | 0.048 | 0.035 | 365 | 1.37 | [-0.021, 0.117] | 0.172 |
| **Mean Cond** | -0.011 | 0.049 | 365 | -0.23 | [-0.107, 0.085] | 0.818 |
| **SD Cond** | 0.175 | 0.052 | 365 | 3.37 | [0.073, 0.277] | < 0.001 |
| **Mean Cond * SD Cond** | -0.060 | 0.069 | 365 | -0.86 | [-0.196, 0.077] | 0.39 |

**Table S10.** (continued)

| *Exp1* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.380 | 0.093 | 323 | 4.06 | [0.195, 0.559] | < 0.001 |
| **Empathy** | -0.000 | 0.001 | 318 | -0.29 | [-0.002, 0.001] | 0.77 |

**Table S11.** **Linear mixed-effects model predicting absolute donation changes from empathy scores, after controlling for confounding variables**

*|D2 — D1|~1 + Empathy Score + Age + Gender + Mean Cond + SD Cond + + Mean Cond * SD Cond + (1 | SubID) + (1 | Charity item)*

| | | | | | | |
|---|---|---|---|---|---|---|
| **Age** | -0.002 | 0.002 | 318 | -0.68 | [-0.006, 0.003] | 0.50 |
| **Gender** | 0.007 | 0.019 | 318 | 0.36 | [-0.030, 0.044] | 0.72 |
| **Mean Cond** | 0.014 | 0.023 | 318 | 0.58 | [-0.032, 0.059] | 0.56 |
| **SD Cond** | 0.002 | 0.024 | 318 | 0.09 | [-0.044, 0.048] | 0.93 |
| **Mean Cond * SD Cond** | 0.030 | 0.033 | 318 | 0.92 | [-0.034, 0.095] | 0.36 |

Note: Gender was coded as non-Female (0) and Female (1).

| *Exp2* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 2.100 | 0.440 | 341 | 4.89 | [1.278, 2.987] | < 0.001 |
| **Empathy** | -0.002 | 0.005 | 338 | -0.47 | [-0.011, 0.007] | 0.64 |
| **Age** | -0.008 | 0.005 | 338 | -1.62 | [-0.017, 0.002] | 0.11 |
| **Gender** | -0.150 | 0.120 | 338 | -1.28 | [-0.391, 0.083] | 0.20 |
| **Mean Cond** | 0.140 | 0.130 | 338 | 1.14 | [-0.103, 0.388] | 0.26 |
| **SD Cond** | 0.093 | 0.130 | 338 | 0.74 | [-0.153, 0.339] | 0.46 |
| **Mean Cond * SD Cond** | -0.034 | 0.180 | 338 | -0.19 | [-0.384, 0.316] | 0.85 |

| *Exp3* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 26.000 | 5.200 | 344 | 5.1 | [16.23, 36.45 ] | < 0.001 |
| **Empathy** | -0.046 | 0.041 | 343 | -1.14 | [-0.126, 0.033] | 0.25 |
| **Age** | -0.330 | 0.160 | 343 | -2.14 | [-0.636, -0.028] | 0.032 |
| **Gender** | 2.300 | 1.100 | 343 | 2.03 | [0.074, 4.509] | 0.043 |
| **Mean Cond** | -1.100 | 1.200 | 343 | -0.94 | [-3.475, 1.218] | 0.35 |
| **SD Cond** | 2.500 | 1.200 | 343 | 2.1 | [0.172, 4.864] | 0.035 |
| **Mean Cond * SD Cond** | -2.900 | 1.700 | 343 | -1.72 | [-6.263, 0.408] | 0.085 |

| *Exp4* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|

**Table S11.** (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| **Intercept** | 14.000 | 4.200 | 367 | 3.24 | [5.347, 21.765] | 0.001 |
| **Empathy** | 0.039 | 0.041 | 367 | 0.95 | [-0.041, 0.119] | 0.34 |
| **Age** | -0.087 | 0.041 | 367 | -2.09 | [-0.168, -0.005] | 0.036 |
| **Gender** | 0.940 | 1.100 | 367 | 0.88 | [-1.157, 3.0400] | 0.38 |
| **Mean Cond** | 0.480 | 1.500 | 367 | 0.32 | [-2.430, 3.386] | 0.75 |
| **SD Cond** | 1.200 | 1.500 | 367 | 0.82 | [-1.665, 4.061] | 0.41 |
| **Mean Cond * SD Cond** | 0.140 | 2.100 | 367 | 0.07 | [-3.985, 4.258] | 0.95 |

**Table S11.** (continued)

| *Exp1* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.595 | 0.163 | 315 | 3.66 | [0.275, 0.915] | < 0.001 |
| **Empathy** | 0.000 | 0.001 | 315 | 0.02 | [-0.003, 0.003] | 0.985 |
| $\alpha$ | 0.039 | 0.085 | 315 | 0.46 | [-0.129, 0.207] | 0.649 |
| $\lambda$ | -0.067 | 0.045 | 315 | -1.48 | [-0.155, 0.022] | 0.139 |
| **Age** | -0.007 | 0.004 | 315 | -1.81 | [-0.015, 0.001] | 0.071 |
| **Gender** | 0.016 | 0.032 | 315 | 0.51 | [-0.047, 0.079] | 0.612 |
| **Mean Cond** | -0.055 | 0.040 | 315 | -1.40 | [-0.133, 0.023] | 0.164 |
| **SD Cond** | 0.049 | 0.042 | 315 | 1.17 | [-0.033, 0.131] | 0.243 |
| **Mean Cond * SD Cond** | -0.008 | 0.056 | 315 | -0.13 | [-0.117, 0.102] | 0.893 |

Note: Gender was coded as non-Female (0) and Female (1).

| *Exp2* | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.843 | 0.139 | 336 | 6.05 | [0.569, 1.118] | < 0.001 |
| **Empathy** | -0.002 | 0.001 | 336 | -1.34 | [-0.005, 0.001] | 0.181 |
| $\alpha$ | -0.323 | 0.096 | 336 | -3.36 | [-0.512, -0.134] | < 0.001 |

**Table S12. Linear regression predicting the social information weight from empathy scores, after controlling for other modeling parameters and confounding variables**

*w ~ 1 + Empathy Score + α + λ + Age + Gender + Mean Cond + SD Cond + + Mean Cond * SD Cond*

| | | | | | | |
|---|---|---|---|---|---|---|
| $\lambda$ | -0.017 | 0.048 | 336 | -0.35 | [-0.111, 0.078] | 0.73 |
| **Age** | -0.003 | 0.001 | 336 | -1.89 | [-0.006, 0.000] | 0.059 |
| **Gender** | -0.020 | 0.038 | 336 | -0.54 | [-0.095, 0.054] | 0.589 |
| **Mean Cond** | -0.036 | 0.039 | 336 | -0.92 | [-0.114, 0.041] | 0.357 |
| **SD Cond** | 0.079 | 0.043 | 336 | 1.87 | [-0.004, 0.163] | 0.062 |
| **Mean Cond * SD Cond** | 0.018 | 0.056 | 336 | 0.33 | [-0.092, 0.129] | 0.744 |

| ***Exp3*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 0.840 | 0.185 | 341 | 4.55 | [0.477, 1.204] | < 0.001 |
| **Empathy** | -0.002 | 0.001 | 341 | -1.44 | [-0.005, 0.001] | 0.151 |
| $\alpha$ | -0.157 | 0.091 | 341 | -1.73 | [-0.337, 0.022] | 0.085 |
| $\lambda$ | -0.078 | 0.051 | 341 | -1.53 | [-0.178, 0.023] | 0.128 |
| **Age** | -0.008 | 0.005 | 341 | -1.57 | [-0.019, 0.002] | 0.118 |
| **Gender** | 0.067 | 0.039 | 341 | 1.69 | [-0.011, 0.144] | 0.091 |
| **Mean Cond** | -0.024 | 0.042 | 341 | -0.57 | [-0.106, 0.059] | 0.57 |
| **SD Cond** | 0.146 | 0.047 | 341 | 3.09 | [0.053, 0.239] | 0.002 |
| **Mean Cond * SD Cond** | -0.058 | 0.059 | 341 | -0.99 | [-0.175, 0.058] | 0.325 |

| ***Exp4*** | ***beta*** | ***SE*** | ***df*** | ***t*** | ***95% CI*** | ***p*** |
|---|---|---|---|---|---|---|
| **Intercept** | 0.496 | 0.145 | 365 | 3.42 | [0.211, 0.780] | < 0.001 |
| **Empathy** | 0.002 | 0.001 | 365 | 1.47 | [-0.001, 0.005] | 0.142 |
| $\alpha$ | -0.410 | 0.094 | 365 | -4.37 | [-0.594, -0.226] | < 0.001 |
| $\lambda$ | -0.040 | 0.051 | 365 | -0.80 | [-0.140, 0.059] | 0.425 |
| **Age** | -0.004 | 0.001 | 365 | -2.59 | [-0.006, -0.001] | 0.01 |

**Table S12.** (continued)

**Table S12.** (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| **Gender** | 0.026 | 0.037 | 365 | 0.71 | [-0.046, 0.098] | 0.476 |
| **Mean Cond** | -0.028 | 0.051 | 365 | -0.55 | [-0.127, 0.071] | 0.581 |
| **SD Cond** | 0.176 | 0.054 | 365 | 3.28 | [0.070, 0.282] | 0.001 |
| **Mean Cond * SD Cond** | -0.035 | 0.072 | 365 | -0.49 | [-0.176, 0.106] | 0.627 |

**Table S13. Linear regression predicting the social information weight from psychopathy scores in the perceptual task, after controlling for confounding variables**

*SocWeight ~ 1 + Psychopathy Score + Age + Gender*

| | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.301 | 0.071 | 365 | 4.21 | [0.160, 0.441] | < 0.001 |
| **Psychopathy** | 0.003 | 0.001 | 365 | 3.10 | [0.001, 0.005] | 0.002 |
| **Age** | -0.001 | 0.001 | 365 | -0.89 | [-0.003, 0.001] | 0.373 |
| **Gender** | -0.002 | 0.026 | 365 | -0.09 | [-0.052, 0.048] | 0.932 |

Note: Gender was coded as non-Female (0) and Female (1).

**Table S14. Linear regression predicting empathy scores from the social information weight in the perceptual task, after controlling for confounding variables**

*SocWeight ~ 1 + Empathy Score + Age + Gender*

| | *beta* | *SE* | *df* | *t* | *95% CI* | *p* |
|---|---|---|---|---|---|---|
| **Intercept** | 0.372 | 0.100 | 365 | 3.72 | [0.175, 0.568] | < 0.001 |
| **Empathy** | 0.001 | 0.001 | 365 | 1.15 | [-0.001, 0.003] | 0.251 |
| **Age** | -0.002 | 0.001 | 365 | -1.57 | [-0.004, 0.000] | 0.117 |
| **Gender** | -0.013 | 0.026 | 365 | -0.50 | [-0.064, 0.038] | 0.621 |

Note: Gender was coded as non-Female (0) and Female (1).

# Data availability

All code and data used to generate results and figures in the paper will be made available upon publication at OSF.


# Acknowledgements

We would like to thank Dr. Vladas Griskevicius for his valuable discussions on experimental design. We also want to acknowledge that the Prolific pilot data collection was supported by a Research Grant from the Council of Graduate Students at the University of Minnesota, awarded to Tao Jin. ChatGPT (GPT-5.3, OpenAI) was used for language editing and wording refinement during the manuscript preparation. All outputs were carefully reviewed by the authors. All scientific content, interpretation and final conclusions were performed solely by the authors.

# Additional information

## Funding

| Funder | Grant reference number | Author |
|---|---|---|
| University of Minnesota (UMN) | Internal Funding | Iris Vilares |
| University of Minnesota (UMN) | Research Grant from the Council of Graduate Students | Tao Jin |



## Author ORCID iDs

**Shen Zhang**: https://orcid.org/0000-0002-5801-286X
**Iris Vilares**: https://orcid.org/0000-0003-1500-3864


# References


**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 https://doi.org/10.1016/j.joep.2015.12.007

**Aknin LB,** Barrington-Leigh CP, Dunn EW, Helliwell JF, Burns J, Biswas-Diener R, Kemeza I, Nyende P, Ashton-James CE, Norton MI (2013) Prosocial spending and *well-being: Cross-cultural evidence for a psychological universal*. *Journal of Personality and Social Psychology* **104**:635-652 https://doi.org/10.1037/a0031578 | PubMed

**Atanassova DV,** Oosterman JM, Diaconescu AO, Mathys C, Madariaga VI, Brazil IA (2025) Exploring when to exploit: The cognitive underpinnings of foraging-type decisions in relation to psychopathy. *Translational Psychiatry* **15**:31 https://doi.org/10.1038/s41398-025-03245-2 | PubMed

**Auten G,** Rudney G (1990) The variability of individual charitable giving in the US. *Voluntas: International Journal of Voluntary and Nonprofit Organizations* **1**:80-97 https://doi.org/10.1007/BF01397439

**Blair R. j. r,** Peschardt K. s, Budhani S, Mitchell D. g. v, Pine D. s (2006) The development of psychopathy. *Journal of Child Psychology and Psychiatry* **47**:262-276 https://doi.org/10.1111/j.1469-7610.2006.01596.x | PubMed

**Brazil KJ,** Cloutier D, De Las Llagas N, McMahon SG, Benevides V, Book AS, Visser BA (2023) The Chameleons of Dating: Psychopathic Traits Are Associated with Mimicking Prosocial Personality Traits in Dating Contexts. *Evolutionary Psychological Science* **9**:260-269 https://doi.org/10.1007/s40806-023-00356-1

**Brazil KJ,** Farrell AH, Boer A, Volk AA (2025) Adolescent psychopathic traits and *adverse environments: Associations with socially adaptive outcomes*. *Development and Psychopathology* **37**:477-489 https://doi.org/10.1017/S0954579424000051 | PubMed

**Brown GDA,** Lewandowsky S, Huang Z (2022) Social sampling and expressed attitudes: Authenticity preference and social extremeness aversion lead to social norm effects and polarization. *Psychological Review* **129**:18-48 https://doi.org/10.1037/rev0000342 | PubMed

**Burke CJ,** Tobler PN, Baddeley M, Schultz W (2010) Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences* **107**:14431-14436 https://doi.org/10.1073/pnas.1003111107 | PubMed

**Camerer C,** Mobbs D (2017) Differences in Behavior and Brain Activity during Hypothetical and Real Choices. *Trends in Cognitive Sciences* **21**:46-56 https://doi.org/10.1016/j.tics.2016.11.001 | PubMed

**Charness G,** Naef M, Sontuoso A (2019) Opportunistic conformism. *Journal of Economic Theory* **180**:100-134 https://doi.org/10.1016/j.jet.2018.12.003

**Charpentier CJ,** Wu Q, Min S, Ding W, Cockburn J, O'Doherty JP (2024) Heterogeneity in strategy use during arbitration between experiential and observational learning. *Nature Communications* **15**:4436 https://doi.org/10.1038/s41467-024-48548-y | PubMed

**Chierchia G,** Piera Pi-Sunyer B, Blakemore S.-J (2020) Prosocial Influence and Opportunistic Conformity in Adolescents and Young Adults. *Psychological Science* **31**:1585-1601 https://doi.org/10.1177/0956797620957625 | PubMed

**Cialdini RB,** Goldstein NJ (2004) Social influence: Compliance and conformity. *Annual Review of Psychology* **55**:591-621 https://doi.org/10.1146/annurev.psych.55.090902.142015 | PubMed

**Contreras-Huerta LS,** Lockwood PL, Bird G, Apps MAJ, Crockett MJ (2022) Prosocial Behavior Is Associated With Transdiagnostic Markers of Affective Sensitivity in Multiple Domains. *Emotion (Washington, D.C.)* **22**:820-835 https://doi.org/10.1037/emo0000813 | PubMed

**Corneille O,** Lush P (2023) Sixty Years After Orne's American Psychologist Article: A Conceptual Framework for Subjective Experiences Elicited by Demand Characteristics. *Personality and Social Psychology Review* **27**:83-101 https://doi.org/10.1177/10888683221104368 | PubMed

**Curtis SR,** Richards DK, Jones DN (2020) The Association between Psychopathy and Influencing Others to Use Substances. *Substance Use & Misuse* **55**:1097-1105 https://doi.org/10.1080/10826084.2020.1729196 | PubMed

**Cutler J,** Wittmann MK, Abdurahman A, Hargitai LD, Drew D, Husain M, Lockwood PL (2021) Ageing is associated with disrupted reinforcement learning whilst learning to help others is preserved. *Nature Communications* **12**:4440 https://doi.org/10.1038/s41467-021-24576-w | PubMed

**da Silva Pinho A,** Céspedes Izquierdo V, Lindström B, van den Bos W (2024) Youths' sensitivity to social media feedback: A computational account. *Science Advances* **10**:eadp8775 https://doi.org/10.1126/sciadv.adp8775 | PubMed

**de Leeuw JR** (2015) jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods* **47**:1-12 https://doi.org/10.3758/s13428-014-0458-y | PubMed

**Díaz-Gutiérrez P,** Boone C, Vyas H, Declerck CH (2024) Neural asymmetry in aligning with generous versus selfish descriptive norms in a charitable donation task. *Scientific Reports* **14**:5793 https://doi.org/10.1038/s41598-024-55688-0 | PubMed

**Dimant E** (2019) Contagion of pro- and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology* **73**:66-88 https://doi.org/10.1016/j.joep.2019.04.009

**Dimant E,** Gelfand M, Hochleitner A, Sonderegger S (2024) Strategic Behavior with Tight, Loose, and Polarized Norms. Management Science. https://doi.org/10.1287/mnsc.2023.01022

**Doyon JK,** Shomstein S, Rosenblau G (2025) Feature identification learning both shapes and is shaped by spatial object-similarity representations. *Communications Psychology* **3**:77 https://doi.org/10.1038/s44271-025-00259-w | PubMed

**Driessen JMA,** van Baar JM, Sanfey AG, Glennon JC, Brazil IA (2021) Moral strategies and psychopathic traits. *Journal of Abnormal Psychology* **130**:550-561 https://doi.org/10.1037/abn0000675 | PubMed

**Ecker UKH,** Sanderson JA, McIlhiney P, Rowsell JJ, Quekett HL, Brown GDA, Lewandowsky S (2022) Combining Refutations and Social Norms Increases Belief Change. *Quarterly Journal of Experimental Psychology (2006)* **76**:1275-1297 https://doi.org/10.1177/17470218221111750

**Faul F,** Erdfelder E, Buchner A, Lang A.-G (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* **41**:1149-1160 https://doi.org/10.3758/BRM.41.4.1149 | PubMed

**Feine G,** Groh ED, von Loessl V, Wetzel H (2023) The double dividend of social *information in charitable giving: Evidence from a framed field experiment*. *Journal of Behavioral and Experimental Economics* **103**:101991 https://doi.org/10.1016/j.socec.2023.101991

**Frey BS,** Meier S (2004) Social Comparisons and Pro-social Behavior: Testing "Conditional Cooperation" in a Field Experiment. *American Economic Review* **94**:1717-1722 https://doi.org/10.1257/0002828043052187

**Frolichs KMM,** Rosenblau G, Korn CW (2022) Incorporating social knowledge structures into computational models. *Nature Communications* **13**:6205 https://doi.org/10.1038/s41467-022-33418-2 | PubMed

**Gaule Anne,** Bevilacqua Leonardo, Molleman Lucas, Roberts Ruth, van Duijvenvoorde Anna C, van den Bos Wouter, McCrory Eamon J, Viding Essi (2022) Social Information Use in Adolescents with Conduct Problems and Varying Levels of Callous-Unemotional Traits. *JCPP Advances* **2**:e12067 https://doi.org/10.1002/jcv2.12067 | PubMed

**Gervais MM,** Kline M, Ludmer M, George R, Manson JH (2013) The strategy of psychopathy: Primary psychopathic traits predict defection on low-value relationships. *Proceedings of the Royal Society B: Biological Sciences* **280**:20122773 https://doi.org/10.1098/rspb.2012.2773 | PubMed

**Gillard ND,** Rogers R (2015) Denial of risk: The effects of positive impression management on risk assessments for psychopathic and nonpsychopathic offenders. *International Journal of Law and Psychiatry, Special Issue on New Challenges in Psychology and Law* **42–43**:106-113 https://doi.org/10.1016/j.ijlp.2015.08.014 | PubMed

**Glenn AL,** Efferson LM, Iyer R, Graham J (2017) Values, Goals, and Motivations Associated with Psychopathy. *Journal of Social and Clinical Psychology* **36**:108-125 https://doi.org/10.1521/jscp.2017.36.2.108

**Goldstein NJ,** Cialdini RB, Griskevicius V (2008) A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research* **35**:472-482 https://doi.org/10.1086/586910

**Gong X,** Brazil IA, Chang LJ, Sanfey AG (2019) Psychopathic traits are related to *diminished guilt aversion and reduced trustworthiness during social decision-making*. *Scientific Reports* **9**:7307 https://doi.org/10.1038/s41598-019-43727-0 | PubMed

**Guay J.-P,** Ruscio J, Knight RA, Hare RD (2007) A taxometric analysis of the latent structure of psychopathy: Evidence for dimensionality. *Journal of Abnormal Psychology* **116**:701-716 https://doi.org/10.1037/0021-843X.116.4.701 | PubMed

**Gunschera LJ,** Brazil IA, Driessen JMA (2022) Social economic decision-making and psychopathy: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews* **143**:104966 https://doi.org/10.1016/j.neubiorev.2022.104966 | PubMed

**Hare RD** (2006) Psychopathy: A Clinical and Forensic Overview. *Psychiatric Clinics of North America, Forensic Psychiatry* **29**:709-724 https://doi.org/10.1016/j.psc.2006.04.007 | PubMed

**Hare RD,** Neumann CS (2008) Psychopathy as a Clinical and Empirical Construct. *Annual Review of Clinical Psychology* **4**:217-246 https://doi.org/10.1146/annurev.clinpsy.3.022806.091452 | PubMed

**Hart W,** Breeden CJ, Richardson K (2019) Differentiating dark personalities on impression management. *Personality and Individual Differences* **147**:58-62 https://doi.org/10.1016/j.paid.2019.04.030

**Iarygina O,** Hornbæk K, Mottelson A (2025) Demand characteristics in human–computer experiments. *International Journal of Human-Computer Studies* **193**:103379 https://doi.org/10.1016/j.ijhcs.2024.103379

**Jin T,** Zhang S, Lockwood P, Vilares I, Wu H, Liu C, Ma Y (2023) Learning whom to *cooperate with: Neurocomputational mechanisms for choosing cooperative partners*. *Cerebral Cortex (New York, N.Y.: 1991)* **33**:4612-4625 https://doi.org/10.1093/cercor/bhac365 | PubMed

**Joiner J,** Piva M, Turrin C, Chang SWC (2017) Social learning through prediction error in the brain. *Npj Science of Learning* **2**:8 https://doi.org/10.1038/s41539-017-0009-2 | PubMed

**Lee H,** Chung D (2022) Characterization of the Core Determinants of Social Influence From a Computational and Cognitive Perspective. *Frontiers in Psychiatry* **13** https://doi.org/10.3389/fpsyt.2022.846535 | PubMed

**Levenson MR,** Kiehl KA, Fitzpatrick CM (1995) Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology* **68**:151-158 https://doi.org/10.1037//0022-3514.68.1.151 | PubMed

**Li R,** Balliet D, Thielmann I, de Vries RE (2025) Revisiting situational strength: Do strong situations restrict variance in behaviors?. *Journal of Personality and Social Psychology* **129**:551-575 https://doi.org/10.1037/pspi0000475 | PubMed

**Lin Y,** Qiao Y, Ma H, Xie B (2025) When psychopathy plays nice: Self-construal *moderate the relationship between psychopathic traits and prosocial behaviors*. *Journal of Research in Personality* **118**:104648 https://doi.org/10.1016/j.jrp.2025.104648

**Liu W,** Pluta A, Charpentier CJ, Rosenblau G (2025) A computational cognitive neuroscience approach for characterizing individual differences in autism: Introduction to Special Issue. *Personality Neuroscience* **8**:e2 https://doi.org/10.1017/pen.2025.2 | PubMed

**Lockwood PL,** Apps MAJ, Valton V, Viding E, Roiser JP (2016) Neurocomputational mechanisms of prosocial learning and links to empathy. *Proceedings of the National Academy of Sciences* **113**:9763-9768 https://doi.org/10.1073/pnas.1603198113 | PubMed

**Lockwood PL,** Klein-Flügge MC (2021) Computational modelling of social cognition and behaviour—A reinforcement learning primer. *Social Cognitive and Affective Neuroscience* **16**:761-771 https://doi.org/10.1093/scan/nsaa040 | PubMed

**Lockwood PL,** Seara-Cardoso A, Viding E (2014) Emotion Regulation Moderates the Association between Empathy and Prosocial Behavior. *PLOS One* **9**:e96555 https://doi.org/10.1371/journal.pone.0096555 | PubMed

**Marsh AA** (2019) The Caring Continuum: Evolved Hormonal and Proximal Mechanisms Explain Prosocial and Antisocial Extremes. *Annual Review of Psychology* **70**:347-371 https://doi.org/10.1146/annurev-psych-010418-103010 | PubMed

**Marsh AA,** Stoycos SA, Brethel-Haurwitz KM, Robinson P, VanMeter JW, Cardinale EM (2014) Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences* **111**:15036-15041 https://doi.org/10.1073/pnas.1408440111 | PubMed

**Molleman L,** Ciranka S, van den Bos W (2022) Social influence in adolescence as a double-edged sword. *Proceedings of the Royal Society B: Biological Sciences* **289**:20220045 https://doi.org/10.1098/rspb.2022.0045 | PubMed

**Molleman L,** Kurvers RHJM, van den Bos W (2019) Unleashing the BEAST: A brief measure of human social information use. *Evolution and Human Behavior* **40**:492-499 https://doi.org/10.1016/j.evolhumbehav.2019.06.005

**Molleman L,** Tump AN, Gradassi A, Herzog S, Jayles B, Kurvers RHJM, van den Bos W (2020) Strategies for integrating disparate social information. *Proceedings of the Royal Society B: Biological Sciences* **287**:20202413 https://doi.org/10.1098/rspb.2020.2413 | PubMed

Najar A, Bonnet E, Bahrami B, Palminteri S (2020) The actions of others act as a pseudo-reward to drive imitation in the context of social reinforcement learning. *PLOS Biology* **18**:e3001028 https://doi.org/10.1371/journal.pbio.3001028 | PubMed

Nakamura JS, Węziak-Bialowolska D, Woodberry RD, Kubzansky LD, Shiba K, Padgett RN, Johnson BR, VanderWeele TJ (2025) Demographic variation in charitable giving and helping across 22 countries in the Global Flourishing Study. *Scientific Reports* **15**:14802 https://doi.org/10.1038/s41598-025-96009-3 | PubMed

Nook EC, Ong DC, Morelli SA, Mitchell JP, Zaki J (2016) Prosocial Conformity: Prosocial Norms Generalize Across Behavior and Empathy. *Personality & Social Psychology Bulletin* **42**:1045-1062 https://doi.org/10.1177/0146167216649932 | PubMed

Olsson A, Knapska E, Lindström B (2020) The neural and computational systems of social learning. *Nature Reviews Neuroscience* **21**:197-212 https://doi.org/10.1038/s41583-020-0276-4 | PubMed

Orne MT (1962) On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist* **17**:776-783 https://doi.org/10.1037/h0043424

Osumi T, Ohira H (2017) Selective Fair Behavior as a Function of Psychopathic Traits in a Subclinical Population. *Frontiers in Psychology* **8** https://doi.org/10.3389/fpsyg.2017.01604 | PubMed

Overgaauw S, Jansen M, Korbee NJ, de Bruijn ERA (2019) Neural Mechanisms Involved in Social Conformity and Psychopathic Traits: Prediction Errors, Reward Processing and Saliency. *Frontiers in Behavioral Neuroscience* **13** https://doi.org/10.3389/fnbeh.2019.00160 | PubMed

Panizza F, Vostroknutov A, Coricelli G (2021) How conformity can lead to polarised social behaviour. *PLOS Computational Biology* **17**:e1009530 https://doi.org/10.1371/journal.pcbi.1009530 | PubMed

Patzelt EH, Hartley CA, Gershman SJ (2018) Computational Phenotyping: Using Models to Understand IndividualDifferences in Personality, Development, and Mental Illness. *Personality Neuroscience* **1**:e18 https://doi.org/10.1017/pen.2018.14 | PubMed

Pfänder J, De Courson B, Mercier H (2025) How wise is the crowd: Can we infer people are accurate and competent merely because they agree with each other?. *Cognition* **255**:106005 https://doi.org/10.1016/j.cognition.2024.106005 | PubMed

Reale KS, Bouchard M, Lim YL, Cook AN, Hart SD (2020) Are Psychopathic Traits Associated with Core Social Networks? An Exploratory Study in University Students. *Social Psychology Quarterly* **83**:423-442 https://doi.org/10.1177/0190272520902105

Reniers RLEP, Corcoran R, Drake R, Shryane NM, Völlm BA (2011) The QCAE: A Questionnaire of Cognitive and Affective Empathy. *Journal of Personality Assessment* **93**:84-95 https://doi.org/10.1080/00223891.2010.528484 | PubMed

Rhoads SA, Gan L, Berluti K, O'Connell K, Cutler J, Lockwood PL, Marsh AA (2025) Neurocomputational basis of learning when choices simultaneously affect both oneself and others. *Nature Communications* **16**:9350 https://doi.org/10.1038/s41467-025-64424-9 | PubMed

Rhoads SA, Gu X, Barnby JM (2024) Advancing computational psychiatry through a social lens. *Nature Mental Health* **2**:1268-1270 https://doi.org/10.1038/s44220-024-00343-w

Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014) Bayesian model selection for group studies—Revisited. *NeuroImage* **84**:971-985 https://doi.org/10.1016/j.neuroimage.2013.08.065 | PubMed

Rilling JK, Glenn AL, Jairam MR, Pagnoni G, Goldsmith DR, Elfenbein HA, Lilienfeld SO (2007) Neural Correlates of Social Cooperation and Non-Cooperation as a Function of Psychopathy. *Biological Psychiatry* **61**:1260-1271 https://doi.org/10.1016/j.biopsych.2006.07.021 | PubMed

Ruff CC, Fehr E (2014) The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience* **15**:549-562 https://doi.org/10.1038/nrn3776 | PubMed

Sakai JT, Raymond KM, McWilliams SK, Mikulich-Gilbertson SK (2019) Testing *helping behavior and its relationship to antisocial personality and psychopathic traits*. *Psychiatry Research* **274**:98-104 https://doi.org/10.1016/j.psychres.2019.02.022 | PubMed

**Schultner DT,** Lindström BR, Cikara M, Amodio DM (2024) Transmission of social bias through observational learning. *Science Advances* **10**:eadk2030 https://doi.org/10.1126/sciadv.adk2030 | PubMed

**Shang J,** Croson R (2009) A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods. *The Economic Journal* **119**:1422-1439 https://doi.org/10.1111/j.1468-0297.2009.02267.x

**Singer T,** Lamm C (2009) The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences* **1156**:81-96 https://doi.org/10.1111/j.1749-6632.2009.04418.x | PubMed

**Singer T,** Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD (2004) Empathy for Pain Involves the Affective but not Sensory Components of Pain. *Science* **303**:1157-1162 https://doi.org/10.1126/science.1093535 | PubMed

**Tatar JR,** Cavanagh C, Cauffman E (2016) The importance of (anti)social influence in serious juvenile offenders with psychopathic traits. *Psychology, Public Policy, and Law* **22**:92-104 https://doi.org/10.1037/law0000074

**Toelch U,** Dolan RJ (2015) Informational and normative influences in conformity from a neurocomputational perspective. *Trends in Cognitive Sciences* **19**:579-589 https://doi.org/10.1016/j.tics.2015.07.007 | PubMed

**Tump AN,** Deffner D, Pleskac TJ, Romanczuk P, M Kurvers RHJ (2024) A Cognitive Computational Approach to Social and Collective Decision-Making. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* **19**:538-551 https://doi.org/10.1177/17456916231186964 | PubMed

**Tusche A,** Böckler A, Kanske P, Trautwein F.-M, Singer T (2016) Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. *Journal of Neuroscience* **36**:4719-4732 https://doi.org/10.1523/JNEUROSCI.3392-15.2016 | PubMed

**Vanags P,** Cutler J, Kosse F, Lockwood PL (2025) Greater income and financial well-being are associated with higher prosocial preferences and behaviors across 76 countries. *PNAS Nexus* **4**:pgae582 https://doi.org/10.1093/pnasnexus/pgae582 | PubMed

**Zhäng L,** Gläscher J (2020) A brain network supporting social influences in human decision-making. *Science Advances* **6**:eabb4159 https://doi.org/10.1126/sciadv.abb4159 | PubMed

**Zhang S,** FeldmanHall O, Hétu S, Otto AR (2025) Advantageous and disadvantageous inequality aversion can be taught through learning of others' preferences. *eLife* **14** https://doi.org/10.7554/eLife.102800.2

**Zhou Y,** Han S, Kang P, Tobler PN, Hein G (2024) The social transmission of empathy *relies on observational reinforcement learning*. *Proceedings of the National Academy of Sciences* **121**:e2313073121 https://doi.org/10.1073/pnas.2313073121 | PubMed

# Peer reviews

## Reviewer #1 (Public review):

This manuscript investigates how people use sequential social information when deciding how much to donate to charity. Across four preregistered experiments, participants first made baseline donations to a set of charities, then observed a sequence of donations from five other people whose mean and variability were experimentally manipulated, and finally made a second donation to the same charities. The authors ask whether the mean and variability of others' donations affect the mean and variability of participants' own donations, and whether individual differences in psychopathy and empathy are associated with responsiveness to social information.

The main behavioral finding is that participants shifted their second donations toward the mean of the donations they observed: generous social information increased donations, whereas stingy social information decreased donations. In contrast, the variability of observed donations had little effect on the mean donation shift, but did affect the variability of participants' subsequent donations, with more consistent social information producing stronger reductions in variability. The authors also fit several computational models and conclude that a hybrid model, in which second donations reflect both participants' initial donations and learned predictions of others' donations, best accounts for the data. Finally, they report that psychopathic traits are positively associated with donation change and with model-derived social-information use, and that this association generalizes to a perceptual social-influence task in Experiment 4.

The paper addresses an interesting question and has several strengths, especially the repeated experimental design, the direct manipulation of social-information statistics, and the attempt to connect descriptive behavior with computational modeling and individual-difference measures. However, several aspects of the design and analysis currently block some of the major conclusions. The behavioral results provide convincing evidence that observed donation levels affect later donation decisions. The current evidence is less decisive for the stronger claims that the winning computational model identifies the underlying mechanism, that individual-level model parameters are robust phenotypes, and that psychopathy specifically increases susceptibility to social information.

Strengths:

A major strength of the manuscript is that it investigates social influence in charitable giving across four preregistered experiments with relatively large samples. The core mean-effect result is replicated across different donation scales, across hypothetical and incentivized settings, and across student and more general online samples. This gives the descriptive behavioral finding substantially more credibility than would be available from a single experiment.

The experimental manipulation is also valuable. Rather than presenting only a single prior donation or a simple group average, the authors expose participants to sequences of donations and independently manipulate the mean and variability of this social information. This design allows the authors to ask not only whether social information changes donation levels, but also whether the distributional structure of that information changes the variability of participants' own responses.

Another strength is the combination of traditional statistical analyses with computational modeling. The hybrid model is a reasonable descriptive candidate because it formalizes the intuitive idea that second donations may depend both on participants' initial preferences and on learned expectations about others' donations. This modeling approach has the potential to clarify mechanisms of social-information use, especially if the validation of the model and its individual-level parameters is strengthened.

Experiment 4 is a sensible extension because it uses an incentivized design, includes a more diverse sample, examines transfer to novel charities, and adds a perceptual social-influence task. These features broaden the empirical scope of the manuscript and make the psychopathy-related findings more interesting, although the perceptual-task result should still be treated as requiring replication.

Weaknesses

The first limitation concerns causal interpretation of the phase effects. Participants always make baseline donations first, then observe social information, and then make second donations to the same charities. There is no non-social repeated-donation control condition. This type of design does support the conclusion that donation changes differ as a function of the observed social-information condition, especially the mean of others' donations.

However, it does not by itself fully isolate social influence from other processes that could also occur between a first and second donation to the same item, such as repeated exposure to the charities, slider familiarity, memory of the first donation, regression to the mean, reduced uncertainty, fatigue, or "the experiment clearly wants me to update" demand effects. This issue is especially relevant for the claim that observing others' donations generally reduces the variability of individual donations. The variability effect may well be socially driven, but the absence of a non-social or irrelevant-information repeated-donation control means that this cannot be decisively demonstrated.

The second limitation concerns the trial-level mixed models. The primary mixed-effects models include random intercepts for participants and items, but do not appear to include random slopes for within-participant or within-item phase effects. Since phase is repeatedly manipulated within participants and items, random-intercept-only models may underestimate uncertainty for some phase interactions, resulting in anti-conservative p-values. The convergent participant-level ANOVA analyses are reassuring, but the trial-level inferential claims would be stronger if the authors reported additional analyses using fuller random-effects structures or other methods that better reflect the repeated-measures structure.

The third limitation concerns model comparison and model validation. The computational models are fit separately to each participant, and model comparison is based on summed information criteria and protected exceedance probabilities derived from those participant-level fits. This is informative about relative conditional fit within the tested sample and model set. However, the manuscript uses the winning model to support broader claims about latent computational mechanisms, individual computational phenotypes, psychopathy-related susceptibility, and potential intervention relevance. For these claims, the relevant prediction target is generalization to new participants, whose individual parameters are not known in advance. The current model-comparison approach is not well aligned with that target. Additionally, the loss appears to combine prediction trials and donation outcomes, so the selected model may more strongly reflect performance at predicting participants' guesses about others rather than specifically predicting their own donation decisions.

The fourth limitation concerns the model adequacy checks and recovery analyses. The analyses described as posterior predictive checks do not appear to be posterior predictive checks, because the models are not Bayesian and there consequently isn't a posterior to check. Instead, the analyses appear closer to some sort of in-sample fitted-value reconstruction checks. Such checks provide limited evidence of model adequacy, especially because the same second-donation data used to estimate individual parameters are then used to assess whether the fitted model reproduces the main behavioral patterns. In addition, the reported model and parameter recovery analyses use extremely favorable response-noise assumptions that are not expected to be met in real data. The analyses establish that the models and parameters are mathematically distinguishable in principle, but they do not establish that the individual-level parameters are reliably recoverable under realistic empirical noise levels to the extent required for the analyses performed in the manuscript.

The fifth limitation concerns the interpretation of the psychopathy results. The association between psychopathic traits and donation change is interesting and appears directionally consistent across experiments. However, the interpretation that psychopathy increases susceptibility to social information is vulnerable to biasing by baseline-distance. The manuscript reports that psychopathy is negatively associated with baseline donations in Experiments 1-3. Participants with lower baseline donations have more room to move toward generous social information, and absolute donation change is partly a function of the distance between the initial donation and the observed social mean for mechanical reasons. Thus, an association between psychopathy and absolute donation change could theoretically arise even if psychopathy does not directly increase social susceptibility.

A sixth limitation is that we could not find the links to the preregistration. The authors state when preregistered hypotheses were or were not supported, but it is unclear how these hypotheses were phrased. Most notably, it is unclear how variance in the observed donation choices was supposed to influence participants. As a side note, it was not quite clear if the variance in the observations was higher or lower across charities, across observed persons, or across both.

Several more minor suggestions can also be made regarding the modelling and the presentation of the task, etc.

https://doi.org/10.7554/eLife.112321.1.sa3

## Reviewer #2 (Public review):

Summary:

This manuscript examines how the statistical properties of others' charitable donations shape subsequent giving using four preregistered experiments and computational modelling. The authors find that both the average level and variability of observed donations influence donation behaviour, and that individual differences in social information use are associated with psychopathic traits.

Strengths:

This is a well-executed paper on the important question of how social information shapes charitable giving. In my view, the combination of preregistered experiments, large sample sizes, computational modelling, and a multi-paradigm approach makes for convincing evidence. The progression across experiments, the use of real donation data rather than deception, the incentivized experiment 4, and the generalization to a second paradigm are all notable strengths. The introduction is clearly written and well-motivated - an enjoyable read. The experimental paradigm is thoughtfully designed, and the methods and supplementary materials are described in considerable detail. The computational modelling provides useful additional insights beyond the behavioural analyses.

As far as I could tell, the manuscript also adheres closely to the preregistrations. The primary hypotheses, experimental designs, exclusion criteria, and key analyses are all consistent with the preregistered plans. Deviations seem to consist of methodological improvements (e.g., mixed-effects models replacing ANOVAs), additional computational and robustness analyses, and therefore strengthen rather than weaken the manuscript. (NB: for transparency, I would appreciate a clearer distinction between preregistered and post hoc analyses, as well as a brief explanation for why some preregistered secondary analyses are no longer reported; see minor comments below).

Overall, I enjoyed reading this paper. I believe it will make a valuable contribution. My comments below are intended to further strengthen an already solid manuscript.

Weaknesses:

(1) The rationale for the social-information phase could be clarified further. Given the research question, I wondered why participants observed the five donations sequentially (and only briefly) rather than simultaneously. In particular, variance is arguably more difficult than the mean to encode and remember, and a sequential presentation may both obscure distributional differences and introduce primacy or recency effects. It would be helpful if the authors could better motivate this design choice, and indicate whether they examined possible order effects.

Relatedly, I felt somewhat uncertain about the purpose of asking participants to predict each donation before observing it. The prediction phase appears to play an important role in the computational model, but its theoretical role is not clearly introduced. Is it intended as a measure of participants' evolving beliefs about the descriptive norm, or primarily as a modelling device? Finally, were these predictions incentivized (e.g., for accuracy), and if not, how should readers interpret them?

(2) I would appreciate having the full experimental materials reproduced in the Supplementary Information. This would make it easier to understand what participants experienced during the task, including what they were told about the "other participants" whose donations they observed.

Minor points:

(1) The interpretations around domain-generality would be strengthened by reporting the association between social information use in the charitable giving task and in the BEAST. Currently, both measures are shown to correlate with psychopathy, but it remains unclear whether individuals who rely strongly on social information in one task also do so in the other. Reporting this correlation (or explaining why it cannot be meaningfully computed) would provide a nice and direct test of a domain-general tendency to use social information.

(2) It would help to explain more explicitly why the standard deviation of donations is theoretically interesting in its own right. The motivation for studying the mean seems immediately intuitive, whereas the motivation for focusing on variability could be elaborated on further in the Introduction.

(3) As I said above, I think the manuscript follows the preregistrations closely. Maybe I missed it, but it seems that prediction accuracy and reaction-time analyses were omitted. It would improve transparency further if the authors would briefly mention the preregistered secondary analyses that are no longer reported (and explain why they were omitted).

https://doi.org/10.7554/eLife.112321.1.sa2

## Reviewer #3 (Public review):

Summary:

In this manuscript, the authors aimed to assess the mechanisms of social influence on charitable giving, particularly by separating the role of donation magnitude and variability in others' donations, and by examining the role of incremental social information in a learning framework. They additionally investigated individual differences in the magnitude effects in relation to self-reported psychopathy and empathy. The main findings suggest that magnitude and variability of others' donation impacted the magnitude and variability of the participants' donations, respectively, and that the weight of social information on individual decisions correlates positively with psychopathy, but not with empathy.

Strengths:

(1) The findings extend previous evidence for social influence on charitable giving to contexts where social information is provided incrementally, and to effects on the variability in social information (in addition to the mean).

(2) Individual differences suggest a role for psychopathy, but not empathy.

(3) Findings are replicated across all 4 (or for some findings 3 out of the 4) experiments, which helps strengthen the claims.

(4) Multiple experiments are a strength, especially Experiment 4, which helped address concerns/potential confounds in the previous experiments, increase representativeness of the sample, add incentive compatibility, and generalize to another task domain (perceptual).

(5) For modelling, strong model and parameter recovery was obtained, thus validating the modelling pipelines.

(6) The experiments were pre-registered, though it's unclear whether only planned analyses were pre-registered, or specific directional hypotheses. It would help if the manuscript took the reader through the pre-registration (and any deviation from it), instead of expecting the reader to do the comparison between the pre-registrations and actual manuscripts.

(7) The studies are appropriately powered, and power analyses are provided.

Weaknesses

(1) Lack of rationale and justification for the between-subjects design.

While this design may be appropriate in some cases (for example, for the generalization of donation to new charities or as a potential "intervention"), it would have been great to know if the findings related to social influence extend to a within-subjects design, especially given the weak results related to the effects of standard deviation in others' donations. It is possible that variability in others' responses would have a stronger effect if manipulated within individuals, since the same individual exposed to both high-SD and low-SD social information may weight low-SD information more, but this effect may lack when individuals are only exposed to the same variability across trials.

(2) Motivation for the RL framework.

The use of reinforcement learning (RL) isn't very well motivated, both in the introduction and methods/results (given the task). In particular, why is RL relevant to studying the problem of social influence, which isn't inherently a learning problem? This should be better motivated in the introduction. Second, when taking the task into account, it's unclear why RL is an appropriate model, given that from the perspective of the participant, the 5 others are different individuals, so the model shouldn't assume that predicting an individual's donation should be related to the previous individual's donation. Unless participants are informed that there is some dependency between the 5 donors they observe on each trial? If so, this should be made clear.

(3) Specifics of modelling analyses, and separability between prediction and second donation data.

Does the RL-based model (either prediction-only or hybrid) explain more variance in second donations than a simple linear regression model predicting second donation from initial donation and the mean of others' donations (or each individual other's donation)? It could be helpful to add some models that include social influence (i.e., integration of social and individual information) but no learning mechanisms per se. If this is not done, I do not believe that current results show that participants combine "their initial self-donation tendencies with their predictions of observed others' giving to guide their second individual donations". While participants may update their predictions, the authors should test multiple models of prediction update (fit only on the prediction data to understand the specific mechanisms of prediction update independently of second donation - for example, is it RL, or could it just be a running average, or some other heuristic? In parallel, it would be helpful to test whether it's the learned predictions (or whatever other prediction update mechanism was found to best explain the prediction data) or the actual others' donation information that best explains second donation - when combined with initial donation. These latter models would be fit on second donation data only in order to be comparable. If it's not possible to separate people's predictions from the actual social information (others' donations) then this should be acknowledged as a limitation. Ultimately, separating the modelling by data type (prediction only vs second donation data only) would help provide more insights into the learning mechanisms (if any) and whether it's learned prediction, or just social information, which influences second donation.

(4) Missing statistics in generalization to novel donation results.

On page 13, in the generalization effect, the authors mention that "Compared with participants exposed to High-SD social information, those exposed to Low-SD social information exhibited less variability in their novel donations, with this effect being especially pronounced in the Low-Mean condition." Was this supported by a significant interaction between SD and Mean condition? If so, please report the statistics of the interaction; if not, it's probably better to refrain from making this claim.

(5) Behavioral index of social influence individual differences.

For the first analysis reported on the association with psychopathy (Figure S9), as well as empathy (Figure S10), the absolute change between first and second donation does not seem like the appropriate marker of social influence. While I understand from Figure 2 that most participants changed their donation in a direction consistent with the social information, it would appear more appropriate to calculate an index of donation change consistent with influence, so calculated as D2 - D1 for the high mean groups and D1 - D2 for the low mean groups. This would be a better measure to interpret high values as an index of social influence.

(6) Interpretation of psychopathy effects.

a) The general idea that high psychopathy would be associated with increased social influence seems counterintuitive. While I appreciate that the authors controlled for additional variables such as age, gender, condition, and other model parameters, is it possible that this effect could be instead explained by the availability heuristic (the social information is more readily available to participants than their individual choice from the baseline trials), lower memory for their own choice, or lower IQ/cognitive abilities? These appear to be important confounds to address to be able to interpret the findings.

b) Related to this, and given that psychopathy/empathy were negatively/positively related to baseline donation amounts, it would be good to account for baseline mean donation amount in the individual difference analyses.

c) Finally, the authors interpret this association in line with other studies that have shown strategic social blending in psychopathy - while this seems possible in contexts where others are present, it doesn't really seem to be the case in this task. Did participants believe the other donors were watching them somehow? It also appears contradictory for the incentivized experiment, whereby if high psychopathy participants would no longer be able to "maintain a favorable social image while still pursuing their own self-interests" (p.23), since as soon as incentivization is added, participants' own self-interests are directly in conflict with the social image. Was participants' understanding of the incentive compatibility tested in Experiment 4?

(7) Asymmetry between generous vs stingy social influence and link with psychopathy.

a) Was such an asymmetry present - in other words, were people more strongly influenced by generous others or stingy others, or were the two comparable? I believe some analyses could be added to test this, and this is also where a within-subject design could help (e.g., different parameters for the two directions of social influence at the individual levels).

b) Related to that, does the correlation with psychopathy vary between conditions? It appears important to test if the increased social susceptibility is general or specific to increases (~high mean group, generous social influence) or decreases (~low mean group, stingy social influence) in donation. I understand that the main effect of psychopathy survived controlling for conditions, but it would still be interesting to test for an interaction between psychopathy and condition in predicting donation changes (calculated as suggested in point 5 above) or social influence weight.

(8) Perceptual task in Experiment 4.

a) While it is good to show that there was no correlation between psychopathy and initial estimate in the perceptual task, were there differences in initial estimate accuracy (i.e., difference between initial estimate and correct answer) along psychopathology? If so, this should be controlled for in the analyses. Given that social influence is always in the direction of the true value, the proportional deviations between initial estimate and social information could yield larger numerical differences and induce larger changes in estimate.

b) Even if previous studies have excluded rounds in which participants update their estimate in the opposite direction of the social information or move beyond it, I believe analyses that include those rounds should be included, especially in the context of individual difference analyses. Could it be that individuals who are high in psychopathy or low in empathy have a higher proportion of rounds where they go against the social influence? The same question applies to the main 4 experiments (in case this criterion was applied to) as well as the perceptual task.

c) Because the perceptual task was completed by the same participants as Experiment 4, were the two social influence measures correlated across tasks? Was psychopathy better predicted by a combination of predictors across the two tasks?

(9) Were individual difference measures examined in relation to the variability effect?

(10) Discussion.

The authors argue against a role for opportunistic conformity. While I tend to agree with their interpretation, I believe that it could be strengthened as follows:

a) First, it relies on a null result (the absence of a difference in decreases between low-mean low-SD and low-mean high-SD groups), which I do not believe was explicitly tested; and even if it was, it should ideally be corroborated by Bayesian statistics to provide strength of evidence for the null effect.

b) Second, this could be a great opportunity to dive into the mechanisms of social influence in the model, by testing the theory that only the lowest (or highest) donation from the group (rather than the mean, or the learned prediction) influences donation. Could a subset of participants be better fitted by such a model?

(11) Methods. Maybe I missed it, but it's unclear what participants were told about the other donors they are observing. It is mentioned that they were fully debriefed after the experiment, but what they were told in the instructions appears important. Was believability tested (this also relates to my comment #1 about the rationale for a between-subjects design, which creates fairly biased sets of social information from the perspective of a single participant)? And related to my comment #2, what participants were told about the donors could help justify the rationale for the RL framework.

https://doi.org/10.7554/eLife.112321.1.sa1

## Author response:

We thank the editors and reviewers for their thoughtful and constructive comments on our manuscript. We are pleased that they considered the core behavioral findings important and robust, especially the results showing that the magnitude and variability of others' donations affected the magnitude and variability of participants' donations, respectively. We also appreciate their acknowledgement of the strengths of the experimental design, large sample sizes, the incentive-compatible and across-domain measures included in Experiment 4, and combined behavioral and computational approaches.

We agree that the manuscript would benefit from greater clarification in several areas, further analyses, and more cautious interpretations. In the revised manuscript, we plan to clarify the rationale for sequentially presenting social information, the role of prediction responses, the theoretical motivation of examining the variability of others' donation, the use of the between-subjects design, and the motivation for the RL framework. We also agree that the lack of a non-social repeated-donation control condition limits the interpretation of the phase effects. Our design permits strong inferences about differences in donation changes across different conditions, but it cannot establish that the phase-related changes are exclusively attributable to social information exposure. We will revise the wording accordingly, moderate the causal language, and explicitly discuss the limitations of our design.

To strengthen the behavioral analyses, we plan to supplement the current mixed-effects linear models with models that reflect the repeated-measure structure of the task, including random slopes for the phase. We will also add statistics in the generalization results section and test the asymmetry between generous vs. stingy social influence. Moreover, the reviewers raised an important concern regarding the associations between psychopathy and the donation change. Because psychopathy is negatively correlated with initial donations in several experiments, absolute donation changes may partly reflect the distance between initial donation and the observed donation mean. We therefore plan to reanalyze the psychopathy effects by using signed donation changes and trial-level discrepancies between participants' initial donations and the observed social information. These additional analyses will enable a more direct and precise assessment of whether psychopathy is associated with greater susceptibility to social influence.

We further agree that the comparison and validation of the computational models should be strengthened. In the revised manuscript, we plan to clarify that the learning models are intended to describe the updating beliefs about a group-level donation norm from sequential social information, rather than learning about a single donor. We will expand the candidate model set to include non-learning models, such as models based on the actual social mean, a running average. We will also model the prediction phase and the second donation phase separately. This will help identify the models that provide explanatory values for both predictions of others' donations and individual donation behaviors. In addition, because the models were not estimated via a Bayesian framework, we agree that the term "posterior predictive checks" is inappropriate. We will rename these analyses. We will also rerun the parameter and model recovery analyses using empirically informed noise levels separately for the prediction and donation phases. In addition, we will implement model-evaluation processes, such as cross-validation, that better reflect prediction for new participants.

In Experiment 4, we plan to directly report the association between social-information-use measures in the perceptual and the donation task to strengthen the domain-generality effect. We will additionally examine whether social susceptibility in the perceptual task is associated with psychopathy by including all trials, including those in which participants moved away from or beyond the social value.

Finally, we will correct the reporting and presentation issues identified by the reviewers, including the social information use equation in the perceptual task, the pseudo-SD of individual donations formula, supplementary figure captions, and task duration. We will also provide fuller experimental materials and make the preregistration links more prominent. In addition, we intend to make the analysis code, model-fitting scripts, and data available during the revision process.

We greatly appreciate the editors' and reviewers' thoughtful suggestions, which will help us substantially strengthen the manuscript. We are grateful for the opportunity to address these important points and believe that the planned revisions will enhance the manuscript's clarity, robustness, and its contribution to the understanding of social influence in donation behaviors.

https://doi.org/10.7554/eLife.112321.1.sa0