The Role of Interpersonal Uncertainty in Prosocial Behavior

Anujit Chakraborty   Luca Henkel

September 17, 2025

Abstract

In prosocial decisions, decision-makers face interpersonal uncertainty–uncertainty about how their choices impact others' utility. We show theoretically and demonstrate empirically how interpersonal uncertainty shapes behavior across key paradigms in the social preference literature. We hypothesize a two-part mechanism: individuals are averse to interpersonal uncertainty, and they experience different interpersonal uncertainty for different people. We use three approaches to show how this mechanism influences prosocial behavior. First, we compare standard allocation decisions, such as dictator games, with decisions where we remove social consequences but retain uncertainty, revealing strikingly similar patterns. Second, we experimentally vary interpersonal uncertainty to disentangle and quantify its contribution relative to social preferences in prosocial decisions, which we estimate to be of similar importance. Finally, we show that self-reported interpersonal uncertainty systematically predicts behavior across individuals, allocation patterns, and interventions that increase charitable giving.

Keywords: prosocial behavior, decision-making under uncertainty, interpersonal uncertainty, ingroup favoritism, self-favoring behavior

JEL Classification: C91, D01, D91

Contact: Anujit Chakraborty: University of California, Davis; chakraborty@ucdavis.edu. Luca Henkel: Erasmus University Rotterdam, ERIM, CESifo, JILAEE, IZA, Tinbergen Institute; henkel@ese.eur.nl.

Acknowledgements: We thank Peter Andre, Björn Bartling, Douglas Bernheim, Felix Chopra, Bikramaditya Datta, Josh Dean, Benjamin Enke, Armin Falk, Simon Gächter, Andrés Gago, Thomas Graeber, Yucheng Liang, John A. List, Ted O'Donoghue, Ryan Oprea, Devin Pope, Chris Roth, George Wu and Florian Zimmermann for valuable comments. Dasha Khamaza provided excellent research assistance.

Funding: Funded by the German Research Foundation (DFG) under Germany's Excellence Strategy - EXC 2126/1- 390838866, UCD Small Research grants (FS24ECH) and Global Affairs Seed grant.

Ethics approval: The study was approved by the ethical committees of the University of Chicago (#24-0099) and the University of California, Davis (#2134317-1).

Research transparency: All studies were preregistered (#159530, #161634, #174673, #159768, #162610, #180119, #209617). See Appendix N for details. The instructions of all studies are available at OSF.

1 Introduction

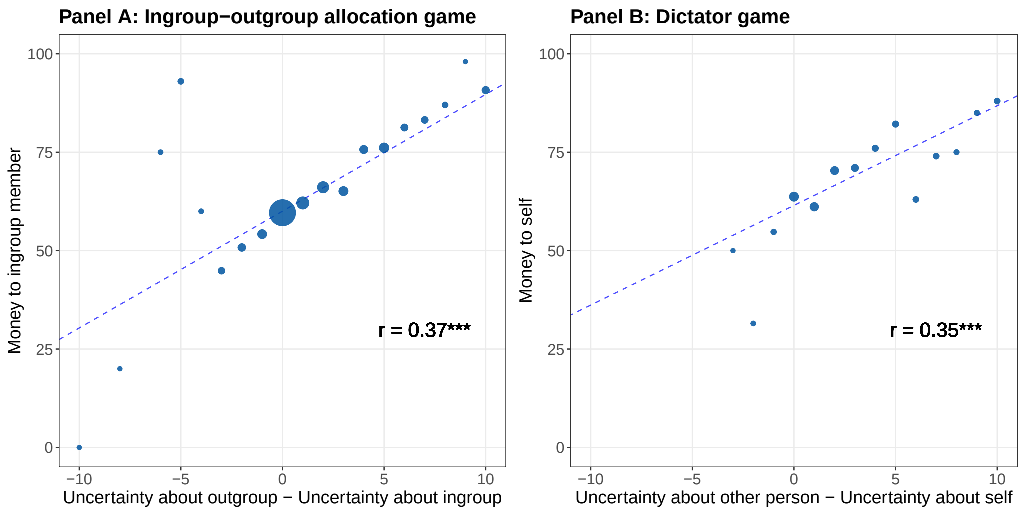
We as humans can only experience our own utility but not others' utility. Thus, prosocial decisions are inherently made under uncertainty, where we are uncertain about how our decisions impact others' utility. We refer to this uncertainty present in prosocial decisions as *interpersonal uncertainty*. If individuals respond to interpersonal uncertainty as they do to other forms of uncertainty, their responses will shape their prosocial decisions. Yet, most theories of prosocial behavior don't model interpersonal uncertainty, and empirical studies typically interpret and estimate prosocial behavior under the assumption of certainty.

In this paper, we show theoretically and demonstrate empirically how interpersonal uncertainty shapes behavior across key paradigms in the social preference literature. We hypothesize a two-part mechanism: (i) individuals are *averse* to interpersonal uncertainty, and (ii) they experience *different* interpersonal uncertainty for different people. Specifically, individuals may experience greater uncertainty about outgroup members than ingroup members, and greater uncertainty about others compared to themselves.

We use three complementary experimental paradigms to provide evidence that our two-part mechanism drives ingroup/outgroup and dictator game behavior. First, we design both *social* decision scenarios –standard allocation tasks that capture ingroup/outgroup and dictator game behavior– and novel *non-social* decision scenarios that remove the scope for social preferences while retaining subjective interpersonal uncertainty. This design allows us to isolate how subjects' perceived uncertainty and their response to it influence behavior in the presence and absence of social preferences. Second, we ask subjects to allocate money between members of two synthetic groups, where interpersonal uncertainty and preference-based influences are independently (and objectively) varied at the group level. This allows us to measure the causal effect of objective interpersonal uncertainty on prosocial behavior, and to quantify subjects' risk attitude toward it, i.e., the first part of our mechanism. Third, we directly elicit the *interpersonal uncertainty* subjects experienced when making decisions in the ingroup-outgroup and dictator game environments. We use these self-reports to quantify the additional uncertainty reported about the outgroup compared to the ingroup, and about others' valuations compared to their own, assessing the second part of our mechanism.

Our results from all three experimental paradigms confirm that interpersonal uncertainty is an important driver of *ingroup favoritism* and *self-favoring behavior*. Conceptually, our findings imply that prosocial behavior and, in particular, charitable giving, is not just a function of stable social preferences, but also of how clearly individuals can imagine or understand the recipient's utility. In this sense, charitable behavior is shaped not only by how much one cares, but also by how much one knows –or *thinks one knows*– about the recipient. As a psychological construct, interpersonal uncertainty can thus be understood as *empathic confidence* –a metacognitive measure of how accurately people believe they can predict others' well-being or utilities.

Figure 1: Association between reported interpersonal uncertainty and allocation decisions



Notes: Binned scatter plots showing the relationship between reported interpersonal uncertainty and decision-making in ingroup versus outgroup allocation games (Panel A) and the dictator game (Panel B). **Panel A:** The x-axis shows the difference in interpersonal uncertainty between ingroup and outgroup members, and the y-axis shows money allocated to the ingroup member. $n = 720$ decisions from 240 subjects, pooled across three social groups and two environments (*Gift card* and *Effort*, see Section 3). **Panel B:** The x-axis shows the difference in interpersonal uncertainty between the dictator game recipient and dictator, and the y-axis shows the amount allocated to the dictator (self). $n = 120$ decisions from 120 subjects. In both panels, we divide the interpersonal uncertainty measure into 20 bins and plot the average allocation in each bin. The blue dotted line indicates the linear regression fit, and the correlation coefficients are displayed (***p<0.01).

Methodologically, our results show that prosocial behavior reflects both social preferences and beliefs in the form of interpersonal uncertainty, implying that any estimation of social preferences can be biased if uncertainty is not accounted for. Our experimental paradigms offer a way to disentangle the effects of social preferences and interpersonal uncertainty, and to identify and quantify their separate contributions. For example, in our synthetic group setting, we find that both factors contribute roughly equally to prosocial behavior. Moreover, our proposed measure of interpersonal uncertainty–elicited directly from subjects–strongly and systematically predicts the extent to which they favor their ingroup or themselves in monetary allocation decisions (see Figure 1).

In a separate set of experiments, we find that interpersonal uncertainty also predicts how prosocial behavior responds to well-known behavioral interventions. Specifically, we implement *identifiable victim effect* and *contact interventions* and show that individuals become more generous when these interventions succeed in reducing their uncertainty about the recipient's valuation or experience. This suggests that a key pathway through which such interventions work is by lowering interpersonal uncertainty.[1] This malleability of interpersonal uncertainty, as measured in our treatments, helps explain why some interventions succeed while others fail, and why generosity often appears context-dependent. Overall, our results suggest that interpersonal uncertainty serves as a unifying mechanism that, alongside social

[1]This interpretation aligns with widespread practices in charitable fundraising—such as using vivid storytelling, photographs, and personal narratives—which reduce uncertainty about their needs and circumstances.

preferences, helps explain prosocial behavior across different paradigms and interventions.

Social and non-social decisions. We illustrate our approach of using *social* and *non-social* decisions with the ingroup versus outgroup paradigm. In the *social* decision, a decision-maker (DM) has to allocate money between two randomly matched individuals. The two individuals receive the allocated money in the form of gift cards: the decision thus has consequences for both. The DM is told that one of the individuals shares similar hobbies/ political views/ religious beliefs, making them an ingroup member, while the other does not, making them an outgroup member. Allocating more money to the ingroup member is typically interpreted as an expression of an explicit *preference* or *taste* for the ingroup. For instance, a DM may get a higher marginal utility from the benefit received by the ingroup member (U_{in}) compared to the outgroup member (U_{out}).

We design the *non-social* decisions to rule out any such preference or taste-based channel but retain the interpersonal uncertainty. As before, the DM splits gift card money between an ingroup and an outgroup member, but now without any consequences for either. Instead, the DMs themselves are paid the sum of their matched ingroup and outgroup members' utilities U_{in} and U_{out} from receiving the gift card money. To do so, we measure U_{in} and U_{out} by eliciting the ingroup and outgroup members' willingness-to-pay (WTP) to receive gift card money. The DMs are paid the sum of the two WTPs, weighted by their allocations to the respective members. Importantly, because both WTPs contribute symmetrically to the DMs' total payment, DMs no longer have any taste-based reason to favor either. However, since DMs do not know the WTPs, they face interpersonal uncertainty, and their uncertainty about the distribution of U_{in} and U_{out} might be asymmetric.

In particular, we hypothesize that the DM perceives higher interpersonal uncertainty about U_{out} than U_{in} , for instance, due to her lower familiarity and fewer interactions with the outgroup, and that uncertainty contributes to her ingroup favoritism. As we derive in our theoretical framework, higher uncertainty about the outgroup is indeed sufficient to generate ingroup favoritism under risk aversion. Intuitively, allocations to the ingroup are a "safer bet" and thus preferable under risk-aversion.

We find that behavior in the *non-social* decisions is similar to the *social* decisions, both in terms of the average and the distribution of allocations. Subjects allocate on average 61% of the endowment to the ingroup member in the *non-social* decisions compared to 63% in the *social*. Leveraging our within design where subjects make both *non-social* and *social* decisions in a randomized order, we find a high correlation of social and non-social decisions at the individual level ($r = 0.53$) with the median subject making identical choices in both decisions. Because interpersonal uncertainty is the only behavioral factor common to both decisions, these results highlight it as a key driver of the ingroup-favoritism pattern.

To ensure that the observed similarity between the *social* and *non-social* decisions is not confounded by subjects being confused or inattentive about the incentives, we run three robustness experiments. In these, we systematically vary the *non-social* incentives. In the first, we increase the multiplier on the ingroup member's WTP, which incentivizes more ingroup giving. In the second, we increase the outgroup member's WTP multiplier, incentivizing more outgroup giving. In the third, subjects are instead paid the minimum of the ingroup and outgroup members' allocation-weighted WTP, which incentivizes equal allocations. We find that subjects strongly respond to these changes: ingroup favoritism increases in the first treatment, flips to outgroup favoritism in the second, and vanishes entirely in the third. These results demonstrate that *non-social* behavior is a conscious and deliberate decision.[2]

Synthetic groups. Our next treatment uses synthetic groups to exogenously vary interpersonal uncertainty, and thus causally identify the aversion towards interpersonal uncertainty. In the treatment, subjects allocate money between individuals randomly chosen from two exogenously created groups. We provide subjects with the actual valuation distribution within each group and exogenously vary this distribution across decisions. Crucially, the greater the variance in valuations within a group, the higher the interpersonal uncertainty about recipients of that group. Independently, we also vary whether the group members share the DM's social group (ingroup status). The factorial variation of group information and interpersonal uncertainty allows us to isolate the marginal influence of each factor.

In the absence of ingroup/outgroup information, subjects allocate on average 60% of their endowment to the group having a lower variance in valuations, directly revealing an aversion to interpersonal uncertainty. When we provide both ingroup information and the distribution of valuations, subjects on average allocate 64% to the ingroup when the outgroup's valuations are more uncertain and 49% when the ingroup's valuations are more uncertain. Lastly, when all members across both groups have the same valuation, and thus interpersonal uncertainty is absent, subjects allocate 57% to the ingroup. Accordingly, changes in interpersonal uncertainty shape differences in allocations in the direction predicted by our framework.

When we regress allocations on the treatment variations, we find that the marginal response to interpersonal uncertainty is similar in magnitude to the marginal effect of ingroup preference, with both increasing allocations by around 7% of the endowment. We also estimate a structural model that quantifies the effect of each factor in isolation: aversion towards interpersonal uncertainty is best fit through a CRRA parameter of 0.37 and the strength of pure ingroup preference through a 7% higher allocation to the ingroup.

Direct measurement of interpersonal uncertainty. Our third and final piece of evidence for the relevance of interpersonal uncertainty in ingroup versus outgroup decisions directly elicits interpersonal uncertainty from subjects. We asked subjects to separately state (using Likert scales) how certain they are about ingroup and outgroup members' WTPs. As pre-

[2]While we see our robustness treatments as a more direct test for the influence of confusion, we also show that our results of the main experiment are not affected if we exclude subjects that fail comprehension checks.

dicted by our framework, subjects perceive significantly higher uncertainty about outgroup members' WTPs compared to ingroup members' WTPs. At the same time, they perceive no difference in average WTPs. Moreover, higher relative outgroup uncertainty significantly predicts stronger ingroup favoritism in the *social* decisions ($r = 0.37$).

Next, we test whether interpersonal uncertainty explains the malleability of prosocial decisions by implementing two interventions—contact and identifiable victim—known to increase prosociality. We replicate their effectiveness in boosting prosocial behavior and find that both interventions simultaneously lower perceived interpersonal uncertainty compared to controls. These results suggest that interpersonal uncertainty may be a key mechanism driving the success of interventions that promote prosocial behavior.

Explaining self-regarding behavior. Building on this three-part evidence of interpersonal uncertainty shaping ingroup-outgroup behavior, we then investigate its importance in shaping self versus other behavior. We again use our *social* and *non-social* approach, creating two treatments that consist of a *Self social* and a *Self non-social* decision scenario. While the former is a standard dictator game where DMs allocate gift card money between themselves and another person, the *Self non-social* decision again has no consequence for others. Instead, the DMs' incentives are to maximize the sum of their own WTP and the other person's WTP. Since both WTPs contribute equally to DMs payments, a self-preference no longer predicts more allocation to the self. However, since subjects only know their own WTP but not others' WTP, aversion to interpersonal uncertainty alone might generate "selfish looking choices".

We find that *Self non-social* behavior resembles behavior in *Self social*. Compared to the 69% of the endowment that subjects allocate to themselves in *Self social*, they allocate 64% to themselves in *Self non-social*, and we cannot reject the equality of the two distributions. The similarity extends to within-subject comparisons: the two decisions are highly correlated ($r = 0.71$), making the *Self non-social* decision one of the strongest predictors of dictator game behavior in the literature. Moreover, our direct measurements of interpersonal uncertainty show that subjects report significantly greater uncertainty about others' WTP than their own. In addition, higher relative uncertainty predicts the degree to which subjects allocate more to themselves.

Related literature. We demonstrate how interpersonal uncertainty influences patterns of prosocial behavior which have been documented across two different strands of the literature. First, a large literature finds evidence that people behave more prosocially towards ingroup members, a finding that is robust across different groups, domains, and methods (such behavior has been labeled ingroup favoritism, parochial altruism or moral universalism, see Charness and Chen, 2020; Shayo, 2020; Enke, 2023, for recent overviews).[3] Second, many studies have documented that in *self versus other* decisions, most people behave

[3] See Iyengar et al. (2019) and Böhm, Rusch, and Baron (2020) for a review of the recent literature on ingroup favoritism in political science and psychology, respectively.

prosocially but tend to make choices that favor themselves more than others (see Capraro, Halpern, and Perc, 2024; Fehr and Charness, 2025, for overviews). We propose a unified, belief-based mechanism that contributes to both patterns of prosocial behavior.[4] Our results show that due to the inherent presence of uncertainty, observed prosocial behavior cannot be interpreted solely as expressions of social preferences even in standard elicitation tasks. In particular, our results imply that these tasks overestimate the extent of ingroup preferences (or taste-based discrimination) and underestimate the degree of altruism in dictator games. We provide a methodology to separately identify and quantify the roles of beliefs and preferences in driving prosocial behavior.

Our paper also sheds new light on the empathy literature. While prior work shows that perceived similarity or *relatability* increases prosocial behavior (Andries et al., 2024), our framework suggests a novel channel for this effect: relatability may enhance generosity not only by increasing other-regarding concerns, but also by boosting confidence in one's ability to infer and predict others' well-being. In this way, *empathic confidence*–as measured by interpersonal uncertainty–provides a new bridge between the role of relatability in prosocial behavior and the metacognitive processes that govern prosocial decisions.

With our subjective uncertainty-based explanation of prosocial behavior, we relate to a recent literature that explains a range of behavioral patterns through people's cognitive response to (subjective) uncertainty. Enke and Graeber (2023) investigate how people's uncertainty over the optimal decision influences choice under risk, belief formation, and forecasts. In the domain of intertemporal decisions, a series of theoretical studies show that risk and time preferences closely intertwine when DMs are uncertain about future consumption (Sozou, 1998; Dasgupta and Maskin, 2005; Halevy, 2008; Chakraborty, Halevy, and Saito, 2020) or future preferences (Amador, Werning, and Angeletos, 2006; Chakraborty, 2021). While this literature focuses on characterizing a logical equivalence between subjective uncertainty and risk or time preference patterns, we study the connection between subjective uncertainty and prosocial behavior both theoretically and empirically. In particular, our non-social treatments allow us to quantify the extent to which subjective uncertainty in the form of interpersonal uncertainty drives standard patterns of social behavior.[5]

[4]Previous belief-based explanations of prosocial behavior have been mainly applied to strategic interactions, such as trust or reciprocity (Berg, Dickhaut, and McCabe, 1995; Fehr and Gächter, 2000).

[5]We thus differ from papers investigating prosocial behavior under experimenter-induced *objective risk* over consequences to study ex-post versus ex-ante fairness (e.g., Brock, Lange, and Ozbay, 2013), the use of risk to act selfishly (Exley, 2016), or, risk-induced morality (Chen and Zhong, 2025). Further, Cappelen et al. (2022) and Cappelen, De Haan, and Tungodden (2024) study redistribution decisions when the source of inequality is uncertain. Moreover, Kappes et al. (2018) vary uncertainty about the wealth level of recipients, finding evidence against the hypothesis that people exploit such uncertainty to license more selfishness.

2 Conceptual framework

Building on our central premise that decision-makers (DMs) experience *interpersonal uncertainty* when making choices that affect others, we introduce a simple "*as if*" model to formalize this idea. In this model, a DM allocates $\$100$ between two recipients and evaluates each outcome using a function U that combines the recipients' utilities symmetrically.[6] The DM represents the interpersonal uncertainty through two probabilistic priors, one for each recipient's potential utilities. Next, we formally define what it means to experience higher interpersonal uncertainty over one of the priors. Then, we derive testable predictions for DM's decisions that maximize the expectation of a risk-averse U based on the DM's priors.

2.1 Assumptions about interpersonal uncertainty

For simplicity, we assume that the DM is probabilistically sophisticated and believes that dollars are valued non-negatively. Interpersonal uncertainty then means that the DM believes the per-dollar valuation of recipient j is distributed as $v_j \sim f_j$, where f_j is a probability distribution with non-negative support contained in $[0, b]$, and F_j is the corresponding CDF. For $x \geq 0$, we define $S_j(x) = \int_0^x F(y)$.

While we do not take a stance on the precise source of this uncertainty, we outline two plausible mechanisms that justify modeling v_j as a random variable from the DM's perspective. First, recipients may differ in their background wealth in ways unknown to the DM. Such *wealth heterogeneity* can affect how much recipients need or value additional money, even if they share the same utility function. Second, even among recipients with similar wealth, unobservable individual characteristics—such as intrinsic preferences[7] or current needs—can cause *preference heterogeneity* and variation in the marginal utility of money. Such unobservable sources of heterogeneity might motivate the DM's subjective uncertainty over v_j, captured by the belief distribution f_j.[8]

Next, we assume that these belief distributions have two key features. First, a DM understands that different recipients might derive different values from the same allocated dollar amount based on their personalities, past experiences, socioeconomic status, or tastes. Thus, distributions over the valuations of others are non-degenerate.

Second, the belief distributions for different recipients systematically differ, depending on the DM's familiarity with the recipients, or the source of the $\$100$ endowment. For instance, suppose one recipient shares their hobbies, religious beliefs, or political interests with the decision-maker (ingroup member) while the other does not (outgroup member). This makes the allocation decision an ingroup versus outgroup tradeoff, which will be our

[6]Symmetry implies that the DM has no bias towards either recipient, which serves as our benchmark.

[7]For example, two equally wealthy recipients might have different utility curvatures.

[8]As detailed in Section 5, we directly measure interpersonal uncertainty and find robust evidence that DMs indeed perceive such uncertainty.

leading example. Facing this tradeoff, a DM might think that shared interests or identity with a recipient is indicative of shared past experiences, economic status, and tastes. As a consequence, DMs may be better at assessing the individual characteristics of their ingroup, thus perceiving lower interpersonal uncertainty about the ingroup relative to the outgroup. Similarly, in situations where DMs themselves are one of the recipients, they may not perfectly know their own preferences[9], but can assess them much better than others, leading to higher uncertainty about others' preferences. In other situations, DMs might think that one recipient is systematically more likely to have higher valuations than another recipient. Formally:

Definition 1. *DMs perceive a higher interpersonal uncertainty for recipient 2 than recipient 1 if $S_1(x) \leq S_2(x)$ for all x and $S_1(y) < S_2(y)$ for some y. DMs perceive a mean-shifted interpersonal uncertainty for recipient 2a compared to recipient 2b, if there exists $c \in \mathbb{R}_{++}$ such that for all x, $F_{2a}(x + c) = F_{2b}(x)$.*

The condition for higher interpersonal uncertainty is best understood as a generalization of "*f_2 is a mean-preserving spread of f_1*" or equivalently "*f_2 is second-order stochastically dominated by f_1*", because the two quoted notions are defined identically with the additional condition that f_1 and f_2 have equal means. We use the concept of higher interpersonal uncertainty to characterize the optimal allocation x^*. In comparison, we use the mean-shift concept to study how x^* changes when the DM's beliefs about a particular recipient's valuation-distribution shifts to the right.

2.2 Choice behavior under interpersonal uncertainty

We investigate the case of unbiased utilitarian preferences, which means the utility the DM receives from allocating $x \in [0, 100]$ to the ingroup and $(100 - x)$ to the outgroup is $u_{UTIL} = v_1 x + v_2(100 - x)$. As v_1, v_2 are random variables, she maximizes expected utility over the potential utilitarian outcomes:

$$EU(x) = E_{v_1 \sim f_1, v_2 \sim f_2} U\left(v_1 x + v_2(100 - x)\right) \tag{1}$$

where $U' > 0$ and $E_{v_i \sim f_i}$ denotes the expectation with respect to f_i.

Given this setup, the optimal allocation depends crucially on the response to uncertainty as characterized by U, and on the belief distributions f_1 and f_2. We will generally assume that $U'' < 0$ which implies that the DM dislikes higher variance over potential utilitarian outcomes. If both f_1 and f_2 are degenerate with different expected values, the DM will allocate 100 to the recipient with the higher expected value.[10] If both distributions are non-

[9]See Chakraborty (2021) and Gabaix and Laibson (2022) for models where DMs are uncertain about their own (future) utility.

[10]In the trivial case of degenerate distributions with equal expected value, the optimal allocation is non-unique, as the DM is indifferent between all possible allocations.

degenerate, Theorem 1 provides the optimal solutions and serves as our prediction for both the *social* and *non-social* decisions we later employ in our experiments.

Theorem 1. *Suppose DM i has unbiased utilitarian preferences and is risk-averse ($U'' < 0$). If f_1 and f_2 are non-degenerate, independent probability distributions, then*
i) Equal division: If $v_1 \stackrel{d}{=} v_2$ (i.e, $f_1 = f_2$) then i's optimal allocation is $x^* = 50$.
ii) Ingroup favoritism: If f_2 is a mean preserving spread of f_1, then i's optimal allocation is $x^* \in (50, 100)$.
iii) Comparative statics over x^*: Suppose the valuations of the two groups are distributed as v_1 and $c + v_2$ for some constant c and independent random variables $v_1 \sim f_1, v_2 \sim f_2$. Under arbitrary CARA preferences[11], or under CRRA coefficient< 1, the optimal allocation satisfies $dx^*/dc \leq 0$.

For the proof, see the Appendix. Part (i) follows from symmetry: a risk-averse DM hedges against interpersonal uncertainty by allocating equally among ex-ante symmetric recipients. (ii) shows that ceteris paribus, if the DM perceives a higher interpersonal uncertainty about one of the recipients, she allocates more to the other recipient. Accordingly, a DM who perceives higher interpersonal uncertainty about the outgroup member will allocate more money to the ingroup, even if they believe that on average, ingroup and outgroup members benefit equally from receiving money. Similarly, a DM who perceives higher interpersonal uncertainty about other's utility than their own utility will allocate more money to themselves, even if they care about others equally and think that on average, everyone benefits equally from money. This motivates our experiments studying the ingroup versus outgroup paradigm and the self versus other paradigm in Sections 3 and 4. Note that in (ii) we use the assumption of equal expected values simply as a benchmark: our key insight is that interpersonal uncertainty can generate ingroup favoritism despite equal expected values.

Finally, part (iii) shows that the DM would decrease the allocation to the ingroup (or the allocation to herself in the dictator game) if her belief about the outgroup's valuation mean-shifts to the right.[12] For example, if a DM perceives mean-shifted interpersonal uncertainty when allocating a recipient's earned money compared to allocating windfall money (thus, perceiving higher c in the former case), then she would keep less for herself (lower x^*) in the former case. While our main focus is on differences in uncertainty, not means, we study an application of mean-shifted interpersonal uncertainty in Appendix Section K.

Will every commonly used welfare criterion deliver the results of Theorem 1 under the right parameters given our assumptions about interpersonal uncertainty? In Appendix D,

[11]For a utility function $U(w)$, the coefficient of absolute risk aversion (ARA) is defined as $r_1(w) = \frac{-U''}{U'}$ and relative risk aversion (RRA) is defined as $r_2(w) = \frac{-wU''}{U'}$. CARA and CRRA imply r_1 and r_2 are constant respectively.

[12]Under extreme risk aversion, when c increases, the marginal return from the states with high v_2 is so low that on the margin, DMs might prefer to allocate more to v_1 to safeguard their utility in the states where v_2 is low.

we show that Rawlsian preferences are insensitive to higher interpersonal uncertainty. We will use this result later in a robustness analysis to show that people respond to our induced incentives in the expected direction.

Relation to preferences. The economic literature generally interprets ingroup and self-favoritism as expressions of preferences, modeled as a higher utility weight applied to ingroup compared to outgroup members (e.g., Tabellini, 2008) or self compared to others (e.g., Capraro, Halpern, and Perc, 2024). However, as our conceptual model demonstrates, aversion to interpersonal uncertainty can generate ingroup favoritism or self-favoring behavior even under equal and symmetric utility weights. Importantly, symmetric utility weights do not imply the absence of social preferences. To the contrary, the described patterns of prosocial behavior emerge precisely because people have social preferences: they care about others' utility, but as this utility is unobserved, they face uncertainty.

3 Ingroup versus outgroup paradigm

We start by studying ingroup versus outgroup decisions before expanding to further prosocial decisions in later sections. Subsections 3.1 and 3.2 describe how we use social/ non-social decisions and synthetic groups respectively to establish the role of interpersonal uncertainty in ingroup versus outgroup decisions.

3.1 Social and non-social decisions

3.1.1 Experimental design

The experimental sessions using the ingroup-outgroup paradigm feature two distinct decision situations: *social* decisions, which are standard allocation tasks commonly used in the literature, and *non-social* decisions, which eliminate social preference motivations from the social decisions while preserving the inherent interpersonal uncertainty about others' utility.

***Ingroup social* decisions.** For the ingroup versus outgroup paradigm, the *Ingroup social* decision is a "bystander" money-allocation game – one of the standard experimental decision tasks used to identify differential attitudes towards ingroup and outgroup members (e.g., Chen and Li, 2009; Enke, Rodríguez-Padilla, and Zimmermann, 2022). First, each subject answers three questions: (i) selecting their top three hobbies from a provided list, (ii) indicating their position on the political spectrum, and (iii) specifying their religious affiliation (if any). Next, they participate in a game that features three individuals, a decision-maker (DM), one individual who shares a social group with the DM (ingroup member), and another individual who is a member of a different group from the DM (outgroup member). The DM is asked to allocate a fixed reward $\$M$ between the ingroup and outgroup member. In total, DMs face three such allocation decisions. Specifically, they allocate the reward between (i) someone who "shares your interests/hobbies" versus "has different interests/hobbies than you", (ii) someone who "shares your political views (e.g., a fellow left-winger, or a fellow right-winger, etc.)" versus someone who "has different political views than you" and (iii) someone who "shares your religious beliefs (e.g., a fellow Christian, or a fellow atheist, etc.)" versus someone who "has different religious beliefs than you".[13] The degree to which DMs allocate more money to the respective ingroup member reveals their degree of ingroup favoritism.[14]

***Ingroup non-social* decisions.** We design and implement a novel decision situation, the *non-social* decision. In this decision, we remove any other-regarding motivations by removing any consequences the DM's decision has to other individuals.[15] Instead, the DM's choice solely determines their own payoff. Specifically, DMs split \$$M$ between an ingroup member's valuation and an outgroup member's valuation, using the same groups as in *social*, and the DM's payoff Π is determined by the following formula:

$$\Pi\left(x_{in}, x_{out}\right) = x_{in} \cdot v_{in}/M + x_{out} \cdot v_{out}/M.$$

The values v_{in} and v_{out} denote the ingroup and outgroup members' valuation for the M reward, as elicited by the researchers, but unknown to the DM. To scale the incentive, the total valuation for the M reward is divided by M, representing an individual's valuation per unit of the reward medium. The DM's choose x_{in}, which denotes the money split in favor of the ingroup member's valuation, and $x_{out} = M - x_{in}$, which denotes the reward split in favor of the outgroup member's valuation.

This payoff function induces utilitarian preferences because we incentivize the DMs to maximize the sum of the valuations, weighted by the allocations. Since DMs do not know the actual valuations of the matched individuals, such interpersonal uncertainty transforms the *social* decision into an uncertain subjective lottery choice. At the same time, because the valuation is elicited over the same reward medium that is distributed in the *Ingroup social* decision, we retain the interpersonal uncertainty inherent in the *Ingroup social* decision. Importantly, the ingroup and outgroup member's valuations enter the utilitarian payoff function symmetrically, so any differences in allocations are driven by differences in uncertainty about the valuations. We can thus use the comparison of the *Ingroup social* and *Ingroup non-social* decision to assess whether responses to interpersonal uncertainty generate ingroup favoritism.

[13]We use the wording of Enke, Rodríguez-Padilla, and Zimmermann (2022).

[14]Particularly, ingroup favoritism is identified independently of the decision-maker's self-interest. Past research has shown that behavior in such bystander allocation games shows a high test-retest correlation, works equally well when posed hypothetically and incentivized, and is highly correlated with related psychological questionnaires (Enke, Rodríguez-Padilla, and Zimmermann, 2022).

[15]This is reminiscent of Oprea (2024) and Enke, Graeber, and Oprea (2024), who construct diagnostic decisions by stripping risk or discounting-based motivations from standard risky and intertemporal tasks to isolate the role of complexity on decision-making under risk and time.

Eliciting valuations. For robustness, the *social* and *non-social* decisions were run using two environments in separate sessions, labeled *Gift card* and *Effort*, that varied what the v_{in} and v_{out} corresponded to.

In the *Gift card* session, subjects allocated and split $M = \$100$ in the *social* and *non-social* tasks, respectively. In the *social* tasks, the allocated money is sent to the ingroup and outgroup member six weeks from the date of the experiment in the form of Amazon gift card money.[16] In the *non-social* tasks, the valuations v_{in} and v_{out} used to incentivize the DMs corresponded to the ingroup and outgroup recipients' respective willingness to pay (WTP_{in} and WTP_{out}) for a \$100 Amazon gift card to be received in six weeks. The elicited WTPs measured subjects' present-day dollar valuation of the delayed gift card money using a standard incentivized multiple-price-list (MPL) procedure.[17]

To understand how the DMs were incentivized using the WTPs in the *non-social* decisions, suppose the DM split \$40 and \$60 in favor of the ingroup and outgroup member, respectively. If the elicited WTP_{in} and WTP_{out} were 80 and 60, then the DM's payoff would be

$$\Pi\left(40, 60\right) = 40 \cdot 80/100 + 60 \cdot 60/100 = 68.$$

Instead, if the WTP_{in} and WTP_{out} were 60 and 80 respectively, the payoff would become 72. Since DMs did not know the exact WTP_{in} and WTP_{out}, they faced interpersonal uncertainty about it.

In the *Effort* session, subjects allocated and split \$$M = 10$ in the *social* and *non-social* tasks, respectively. The *Non-social* decision of the *Effort* environment closely mimics the non-social decision of the *Gift card* environment, with one difference: here, the valuations v_{in} and v_{out} correspond to the recipient's willingness-to-work (WTW) for the \$10 bonus payment.

To elicit WTW for the ingroup and outgroup members, each recipient faced a series of binary choices between (i) completing a required number of real effort tasks to receive the \$10 bonus and (ii) opting out. Each real effort task required moving 30 sliders to the middle position within a 60-second time limit, following the method of Gill and Prowse (2012, 2019). Similar to the WTP elicitation, the WTW was presented in a multiple-price-list format, with the required number of completed tasks ranging from 0 to 30 in increments of 2. A recipient's WTW was defined as the number of tasks at which they first switched from preferring to complete the tasks to opting out.

Minimizing inattention and confusion. We employ several measures to mitigate inattention and confusion. First, decision-makers complete the relevant WTP or WTW valuation

[16]Amazon gift card money was used to ensure that subjects' valuations differed meaningfully from the dollar value of the gift card. The six-week delay was used to introduce additional variation in subjects' valuations through heterogeneity in impatience.

[17]Subjects made a series of binary decisions between (i) receiving a \$100 Amazon gift card in six weeks and (ii) receiving a monetary amount today, which increased across decisions. We enforced single switching by automatically filling the list above and below subjects' choices.

task themselves before completing the *non-social* decisions. This familiarizes them with the valuation process that underpins the incentive. Second, we include several comprehension questions that test whether DMs understood that the *non-social* decisions only have consequences for themselves, not for the other individuals. If they answer incorrectly on the first attempt, they are informed of the mistake and asked to try again. If they still err, we explain the correct answers and clarify the key concepts. This process reinforces the distinction between *social* and *non-social* decisions. Third, to further minimize inattention, we include an explicit disclaimer on the *non-social* decision screens that states

> "Reminder: your choice only determines your own payment, it does not affect the two individuals."

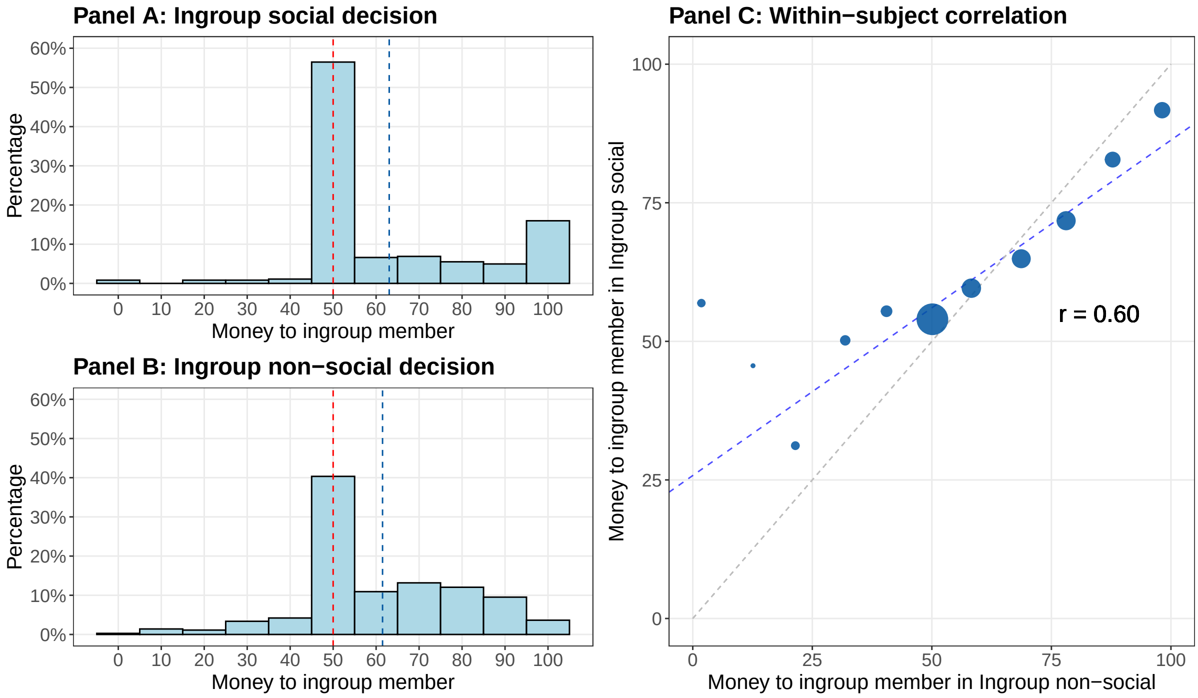
On the decision screen, we also provided DMs with the option to revisit the instructions.

Procedure. We first ran an experimental session in the *Gift card* environment, and then afterwards a session using the *Effort* environment. Within each session, we randomized the order of decisions. Half of the decision-makers first face the *social* decision and then the valuation task and *non-social* decision. The other half first face the valuation task and *non-social* decision, and subsequently the *social* decision. We did not announce beforehand that other decisions would follow the initial decisions, therefore minimizing the scope for contagion from one treatment to the other. This design allows us to analyze within-subject behavior, and compare behavior between-subject by only looking at the first set of decisions.

Data. In total, 240 subjects participated in the ingroup experiment, with 121 subjects first facing the *Ingroup social* decision and 119 subjects first facing the *Ingroup non-social* decision. We ran the experimental session with the *Gift card* environment with 119 subjects, while 121 subjects faced the *Effort* environment. For this and all further experiments, we used Prolific to recruit online participants living in the US. All experiments were preregistered, see Appendix N for details. We used oTree (Chen, Schonger, and Wickens, 2016) for programming the graphical user interface. Subjects spent a median of 10 to 12 minutes in the experiments and received as compensation the equivalent of an hourly wage between \$10 and \$12 per hour. In the *Effort* environment (and subsequent experiments that use WTWs), one out of every ten subjects had one of their decisions implemented with real consequences (between-subject random incentivized system).[18] In the *Gift card* environment (and subsequent experiments that use WTPs), one subject out of all participating subjects had one of their decisions implemented.

[18] In a meta-analysis, Umer (2023) shows that in the context of dictator games paying a subset instead of all subjects does not significantly change behavior.

Figure 2: Main results ingroup versus outgroup decisions



Notes: **Panel A and B:** Histogram of *Ingroup social* (Panel A) and *Ingroup non-social* (Panel B) decisions. The binwidth is 10. In Panel A, the x-axis denotes the amount of money (out of $100) allocated to the ingroup member instead of the outgroup member, with decisions having consequences for the ingroup and outgroup members. In Panel B, the x-axis denotes the amount (out of $100) allocated to the ingroup member's valuation instead of the outgroup member, where the decisions have consequences only for the decision-makers, with their payoff depending on the ingroup and outgroup members' valuation of money. The red line denotes the even split, the blue line the average allocation. **Panel C:** Binned scatter plot of *Ingroup social* and *Ingroup non-social* decisions. We divide *Ingroup non-social* choices into 10 bins and plot the average giving to the ingroup member in *Ingroup social* for each bin. The blue line displays the linear fit of regressing *Ingroup social* on *Ingroup non-social* decisions. The correlation coefficient is $r = 0.60$. Decisions are pooled across the three groups (shared hobbies/interests, political views, and religious beliefs), displaying $n = 363$ (Panel A), $n = 357$ (Panel B), and $n = 720$ (Panel C) decisions by 121, 119, and 240 subjects, respectively.

3.1.2 Results

We start with the between-subject comparison of the *social* and *non-social* decisions. As the results for the *Gift card* and *Effort* environments are quantitatively almost identical, we pool across both.[19] Appendix C presents the results from the two environments separately.

***Ingroup social* decisions.** In the *Ingroup social* decisions, subjects allocate on average $57.75, $70.66, and $60.73 to the ingroup member when they share the same interests/hobbies, the same political views, or the same religious beliefs, respectively. In all three cases, we can reject the hypothesis of no ingroup-favoritism ($p < 0.001$, one-sample Wilcoxon tests). Figure 2 panel A displays the distribution pooled over the three decisions, which replicates the typical distributional pattern found in the literature (e.g., Enke, Rodríguez-Padilla, and Zimmermann, 2022). In 44% of the decisions, subjects display ingroup-favoritism by allocating

[19] We multiply the *Effort* decisions by 10 to ensure a common scale with the *Gift card* decisions.

strictly more than 50% to the ingroup. Outgroup-favoritism is found in 5% of decisions, and in the remaining 51%, subjects allocate 50:50. In total, 71% of subjects display ingroup-favoritism in at least one decision.

***Ingroup non-social* decisions.** Importantly, a similar pattern emerges in the *Ingroup non-social* decisions. Here, subjects allocate on average \$58.61, \$64.60, and \$61.29 in favor of the valuation of the ingroup member that shares their interests/hobbies, political views, and religious beliefs. As before, we find behavior resembling ingroup-favoritism in all three cases ($p < 0.001$, one-sample Wilcoxon tests), even though the decisions have no consequences for ingroup or outgroup members. See Panel B of Figure 2 for the distribution of the pooled decisions. In 59% of the *Ingroup non-social* decisions, subjects display ingroup-favoritism by allocating strictly more than 50% to the ingroup-WTP. Outgroup-favoritism is observed in 12% of decisions, and in the remaining 29%, subjects allocate 50:50. In total, 78% of subjects display ingroup-favoritism in at least one decision.

Comparing *Ingroup social* and *non-social*. We cannot reject that average ingroup allocations are equal between *Ingroup social* and *non-social* decisions at the 5% level in any of the three cases ($p = 0.15$ for hobbies/interests, $p = 0.06$ for political views, $p = 0.15$ for religious beliefs, unpaired Wilcoxon tests). Thus, our *non-social* setup in which decisions have no consequences for either group member produces similar average ingroup favoritism as the standard *social* setup.

When comparing distributions, the most notable difference is the extent of maximal ingroup favoritism, i.e., giving the entire endowment to the ingroup member. In the *social* decisions, 15% of participants display maximal ingroup favoritism, compared to only 3% in the *non-social* condition. This results in significantly different distributions ($p = 0.01$ for hobbies/interests, $p = 0.001$ for political views, $p = 0.01$ for religious beliefs; Kolmogorov-Smirnov tests), suggesting that behavior and in particular extreme ingroup favoritism in *Ingroup social* decisions is jointly shaped by preferences and beliefs –a point we analyze in detail in Section 3.2.1.

Within-subject comparison. Next, we compare behavior between *Ingroup social* and *non-social* on the individual level by including also the second set of decisions of each subject. We replicate the previously reported between-subjects results also within-subject, see Appendix E.1. Importantly, we find no evidence for order effects, supporting the validity of our within-results (see Appendix F for details). This allows us to correlate behavior in *Ingroup social* with *non-social*. Panel C of Figure 2 displays the distribution of each individual social and non-social decision pair in a binscatter-plot. As the figure shows, the two are significantly related: ingroup favoritism in *Ingroup non-social* predicts ingroup favoritism in *Ingroup social*, with a correlation coefficient of $r = 0.60$ ($p < 0.001$).[20] Therefore, the same subjects that display ingroup favoritism when decisions have consequences for others also display it when their decisions solely affect their own payoff, with the payoff depending on others' WTPs.

Result 1. *We find ingroup-favoritism in Ingroup non-social decisions, which retains interpersonal uncertainty but removes any consequences for ingroup or outgroup members. Moreover, decisions are strongly correlated on the individual level with Ingroup social decisions, which feature consequences.*

3.1.3 Potential confounds and robustness treatments

Our main results show a high degree of similarity between the *Ingroup social* and *Ingroup non-social* decisions. However, similar behavior across social and non-social choices may also arise from the following mechanisms that are unrelated to interpersonal uncertainty:

- *Inattention or confusion:* Subjects might misunderstand the incentive structure or fail to fully attend to the distinctions between *social* and *non-social* conditions, leading to superficially similar behavior across both.
- *Indifference:* Subjects may not perceive any interpersonal uncertainty and further believe that ingroup and outgroup valuations are identical, rendering them indifferent between all *non-social* allocations. In such cases, they might replicate their *social* choices as a tiebreaking strategy.
- *Signaling:* Ingroup favoritism in both *social* and *non-social* choices might be driven by the desire to signal concern or loyalty for the ingroup.

To test if inattention or confusion falsely drives our baseline results, we sort subjects based on their responses to the comprehension questions, which also quizzed subjects about the *non-social* incentives. Pooling decisions across the three social groups, subjects with at most one error allocate \$62.32 in *Ingroup social* and \$60.90 in *Ingroup non-social* to the ingroup member. Subjects with more than one error allocate \$64.07 and \$62.50, respectively. The averages remain significantly different from the 50:50 benchmark, in all of the three following cases: for the pooled data, for each social group individually, and when considering other sample splits based on the comprehension performance (see Appendix Figure B.1).

To further test for confounding explanations, we implement three variants of the *Ingroup non-social* condition: the *Ingroup incentive*, *Outgroup incentive*, and *Non-social minimum* treatments.

[20] The correlation is not driven by subjects who display no or maximal ingroup favoritism, as excluding both types of decisions (or either) yields correlations of similar or even stronger magnitude.

Design. The first two treatments vary the weights assigned to the ingroup and outgroup members' WTPs in the decision-maker's payoff function. In the *Outgroup incentive* condition, the weight on the outgroup member is tripled:

$$\Pi\left(x_{in}, x_{out}\right) = x_{in} \cdot WTP_{in}/100 + 3 \cdot x_{out} \cdot WTP_{out}/100$$

Conversely, in the *Ingroup incentive* condition, the weight on the ingroup member is tripled:

$$\Pi\left(x_{in}, x_{out}\right) = 3 \cdot x_{in} \cdot WTP_{in}/100 + x_{out} \cdot WTP_{out}/100$$

In the third treatment – *Non-social minimum* – we replace the utilitarian objective with a Rawlsian one. Here, subjects' payoffs are calculated as:

$$\Pi\left(x_{in}, x_{out}\right) = \min\{x_{in} \cdot WTP_{in}/100, x_{out} \cdot WTP_{out}/100\}$$

Thus, we incentivize subjects to choose the allocation that maximizes the utility of the worse-off recipient, irrespective of group affiliation, and thus under our framework (Section 2), eliminates favoritism in either direction.[21]

Other than the incentives, all aspects of the decisions in the three treatments are identical to the *Ingroup non-social* decisions.

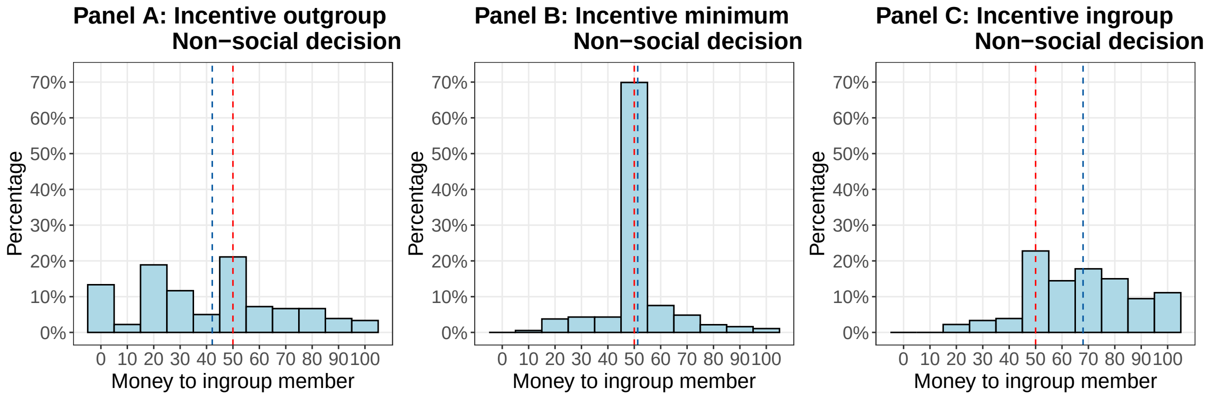
Predictions. Under our model – utilitarianism subject to interpersonal uncertainty – the predictions are straightforward: *Outgroup incentive* should increase allocations towards the outgroup-WTP. *Ingroup incentive* should increase allocations towards the ingroup-WTP. *Non-social minimum* should lead to 50:50 allocations.

Alternative mechanisms make contrasting predictions: Subjects who fail to understand the incentives due to *Inattention or confusion* are unlikely to systematically respond to them and should continue exhibiting default ingroup favoritism across all conditions. For subjects who have no uncertainty and think that outgroup-WTP and ingroup-WTP are identical (*Indifference*), allocating the entire endowment of 100 to the outgroup and ingroup WTPs, respectively constitutes a strictly dominant action in the first two treatments. Subjects desiring to appear loyal to the ingroup (*Signaling*) should not respond to the incentives, in particular not to the *Outgroup incentive* and *Non-social minimum* incentives.

Results. Figure 3 presents the results. Relative to an average allocation of $60.89 to the ingroup-WTP across our social groups in *Ingroup non-social* (Section 3.1.2), we find significant shifts in behavior across all three incentive treatments: subjects allocate on average $42.16 to the ingroup-WTP in *Outgroup incentive*, $51.31 in *Ingroup non-social minimum*, and $67.98 in *Ingroup incentive*, each time a significant difference ($p < 0.001$, unpaired Wilcoxon test). These results are in line with our theoretical predictions based on interpersonal uncertainty: favoritism reverses in *Outgroup incentive*, intensifies in *Ingroup incentive*,

[21] As we show in Appendix D, even if the WTP distribution for the outgroup is a mean-preserving spread of the ingroup's WTP distribution and subjects are risk-averse, the predicted optimal choice under Rawlsian preferences and given our assumptions on interpersonal uncertainty is $x_{in} = x_{out} = 50$, implying no favoritism in either direction.

Figure 3: Incentive ingroup and Incentive outgroup robustness treatment results



Notes: Histogram of *Incentive outgroup* (Panel A), *Non-social minimum* (Panel B), and *Incentive ingroup* (Panel C) decisions, where subjects maximize $(x_{in} \cdot WTP_{in}/100 + 3 \cdot x_{out} \cdot WTP_{out}/100)$, $\min\{x_{in} \cdot WTP_{in}/100, x_{out} \cdot WTP_{out}/100$, and $(3 \cdot x_{in} \cdot WTP_{in}/100 + x_{out} \cdot WTP_{out}/100)$, respectively. The x-axis denotes the amount of gift card money (out of $100) that subjects allocate to the ingroup member's WTP instead of the outgroup member's WTP. The red dotted line denotes the even split benchmark, the blue dotted line the average allocation. The binwidth is 10. Decisions are pooled across the three groups (shared hobbies/interests, political views, and religious beliefs). Panels A and C display $n = 180$ decisions by 60 subjects, Panel B $n = 186$ decisions by 62 subjects.

and disappears in *Non-social minimum*.[22] Not only do averages shift, but the full distributions of choices are also reshaped. For instance, in *Ingroup non-social minimum*, the distribution becomes symmetric around the 50:50 split, while in *Outgroup incentive* the majority of subjects allocate more to the outgroup-WTP. For further details, see Appendix G.

These results offer clear evidence that subjects are sensitive to the incentives in the *non-social* domain: altering a single coefficient in the payoff function flips behavior from ingroup to outgroup favoritism, while implementing a Rawlsian incentive causes favoritism to vanish entirely. This rules out limited attention or confusion as plausible explanations for the ingroup favoritism observed in the *Ingroup non-social* condition. In fact, our robustness treatments allow us to zoom in on participants likely to be inattentive or confused–estimated to be a small minority–and even among them, ingroup favoritism is, if anything, weaker than in the main sample (see Appendix G for details).

The data similarly contradict the *Indifference* account: in nearly 80% of cases, subjects in both *Ingroup incentive* and *Outgroup incentive* choose interior allocations, inconsistent with strict dominance logic under indifference. Finally, the observation of outgroup favoritism–both in average and modal choices–in *Outgroup incentive*, along with the disappearance of favoritism in *Non-social minimum*, undermines the *Signaling* hypothesis.

[22]We find a significantly lower average allocation to the ingroup WTP relative to the 50:50 split in *Outgroup incentive* ($p < 0.001$, one-sample Wilcoxon test) and we can no longer reject that the average is different from the 50:50 split in *Ingroup non-social minimum* ($p = 0.15$, one-sample Wilcoxon test).

3.2 Synthetic groups

According to our conceptual framework from Section 2, even if ingroup and outgroup members are perceived to have equal valuations on average, interpersonal uncertainty drives ingroup favoritism through a two-part mechanism: (1) subjects perceive greater uncertainty for the outgroup than the ingroup, and (2) they are averse to higher uncertainty. In Section 5, we directly measure interpersonal uncertainty to test and validate the first part of the mechanism. In this section, we assess the second part and measure subjects' risk attitudes toward interpersonal uncertainty using synthetic groups that objectively induce interpersonal uncertainty.

3.2.1 Experimental design

Decision. In a new treatment (*Ingroup uncertainty*) with 120 subjects, DMs faced seven randomly ordered allocation decisions. In each decision, DMs were endowed with $10 to allocate between Group A and Group B. Each group consisted of two recipients who were participants of an earlier study in which their willingness-to-work (WTW) was elicited. Recall that the WTW measure captures the number of tasks (ranging from 0 to 30) an individual is willing to complete to earn a $10 bonus payment.

DMs were informed that after their allocation decision, a randomly chosen recipient from each group would receive the money allocated to that respective group. Between decisions, we systematically varied (i) uncertainty over the recipients' WTW within each group and (ii) their ingroup/ outgroup affiliation. This yielded four different decision types where the DMs were *informed* about the following WTW/ group information:

1. *Uncertainty without group information* decision: the two recipients of Group A have the same WTW of 12, while the recipients of Group B have a WTW of 4 and 22, respectively. Thus, while the recipient who eventually gets the money from Group A would have a fixed WTW of 12, the recipient receiving the money from Group B could have a WTW 4 or 22, and DMs knew this. Hence, Group B has a higher variation in WTWs than Group B.
2. *Group information without uncertainty* decision ($\times 2$)[23]: the two recipients of Group A are ingroup members, while the two recipients of Group B are outgroup members. Moreover, all four recipients have the same WTW of 12. Hence, there is no variation in WTWs.
3. *High uncertainty on ingroup members* decision ($\times 2$): the two recipients of Group A are ingroup members, one having a WTW of 4 and the other of 22. The two recipients of Group B are outgroup members, both having a WTW of 12. Hence, the ingroup has a

[23]The "$\times 2$" indicates that there were two decisions of this type, one involving shared hobbies and the other involving shared political views as group identity. We excluded religious beliefs because we aimed to minimize additional decisions, and previous results suggested similar behavior.

higher variation in WTWs.

4. *High uncertainty on outgroup members* decision ($\times 2$): the two recipients of Group A are ingroup members, both having a WTW of 12. The two recipients of Group B are outgroup members, one having a WTW of 12 and the other of 22. Hence, the outgroup has the higher variation in WTWs.

Interpretation. Under the assumption that, ceteris paribus, individuals with higher marginal utility for money exhibit greater willingness to complete tedious and time-consuming tasks to earn it, the WTW serves as a proxy for marginal utility—measured on an effort scale.[24] To validate that Prolific participants interpret WTW differences as indicators of how much individuals value money, we asked participants of our earlier study about their interpretation in an open-ended question. About 73% of the 121 respondents attributed the WTW differences to varying levels of need for money or value of time. For details, see Appendix Section H.

Accordingly, a higher spread in WTWs implies higher uncertainty about recipients' marginal utility from received money – hence higher interpersonal uncertainty. Consequently, the *uncertainty without group information* treatment reveals DMs' attitude towards higher uncertainty about WTWs without the confound of ingroup preferences. Because the expected value of WTWs is higher in Group B than in Group A[25], that is $(22 + 4)/2 = 13 > 12$, DMs who still allocate more to Group A reveal a strict preference against uncertainty in WTWs, and hence their risk aversion. The *group information without uncertainty* decision, on the other hand, reveals ingroup favoritism in the absence of (WTW-based) interpersonal uncertainty. The last two decision-situations reveal the extent to which interpersonal uncertainty influences ingroup favoritism. When presenting the WTW and social group information, we randomized the group's position on the screen (left or right) and which information was presented first, balancing the presentation of the two pieces of information.

Importantly, DMs knew that WTW reflects recipients' *stated* willingness, not their actual exerted effort. This feature ensured that fairness considerations based on exerted effort did not influence their decisions.[26]

[24]DMs in the *Ingroup uncertainty* treatment were familiar with the WTW measurement, as they first completed an example slider task and then underwent the WTW elicitation themselves.

[25]Intrinsically, this argument holds when recipients' value of \$10 is directly equal to their WTWs, as well as when recipients' valuation is equal to an increasing and convex disutility function of the WTW. The latter assumption is well supported by empirical evidence on real effort tasks (see e.g., Gill and Prowse, 2019).

[26]Recipients knew there was a 50% chance they would complete their selected WTW tasks and a 50% chance they would not work but might later receive money from DMs who observed their WTW. Only the latter group was matched with DMs in the *Ingroup Uncertainty* experiment. Thus, none had worked based on their WTW choices—a fact clearly communicated to DMs.

3.2.2 Results

Figure 4 displays the distribution of choices. Starting with Panel A, we observe that the majority of subjects are risk-averse towards interpersonal uncertainty even in the absence of group information: 54% allocate more than 50:50 to the group with lower WTW variance, 27% allocate 50:50 and a minority of 19% allocate more to the group with the higher WTW variance. On average, subjects allocate \$6.00 to the group with lower WTW variance, which is significantly more than the 50:50 benchmark ($p < 0.001$, one-sample Wilcoxon tests).

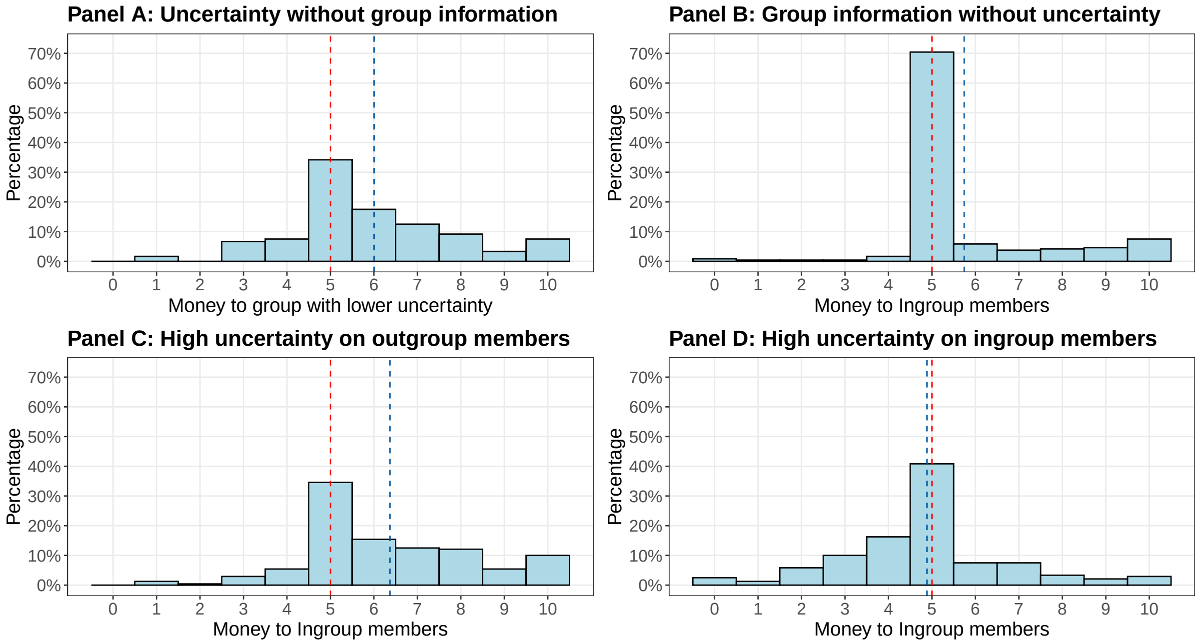
In the case where all recipients have the same WTW (Panel B), subjects allocate on average \$5.74 to the ingroup (pooling the decision across both social groups), which is significantly different from the 50:50 split ($p < 0.01$, one-sample Wilcoxon test). This quantifies the extent to which ingroup preferences drive ingroup favoritism since interpersonal uncertainty about WTWs is absent in this decision.

Panels C to D then document how interpersonal uncertainty being higher in the ingroup or the outgroup changes the magnitude of ingroup favoritism. The amount allocated to the ingroup increases from \$5.74 to \$6.37 when the outgroup is more uncertain (has higher variation in WTWs), and decreases to \$4.89 when the ingroup is more uncertain (both $p < 0.001$ compared to the no uncertainty decision, paired Wilcoxon tests). Moreover, changing the uncertainty changes the entire distribution of choices. When the outgroup has the higher WTW variance, the modal choice (62%) displays ingroup favoritism, while 12% of choices favor the outgroup and 26% are the 50:50 split. Once the recipients of both groups have equal WTWs, the modal choice (66%) is the 50:50 split. Here, in only a minority of 29% and 5% choices is the ingroup and outgroup favored, respectively. Lastly, when the ingroup has the higher WTW variance, the modal choice (40%) switches to outgroup favoritism, while 31% of choices show ingroup favoritism and 29% are 50:50 choices. These results establish that responses to interpersonal uncertainty causally influence the extent of ingroup favoritism.

Information choice. After the seven allocation decisions, we included an eighth decision where DMs, without knowing group membership or WTW, could choose to learn about one of those before allocating money. This decision reveals whether group membership or recipients' valuation primarily drives DM's choices. Overall, 81% of subjects chose to learn about WTW variation, indicating a preference for information about uncertainty over social group membership. For details, see Appendix I.

WTW-based similarity. One concern is that DMs might base their allocation decisions on how similar/ dissimilar the recipients' WTW is to their own, favoring recipients with more similar WTWs. For instance, a DM with a WTW close to 12 might prefer Group A (whose WTW is also 12) simply due to this similarity. While such behavior does not threaten the identification of a causal effect, it would confound our interpretation of the role of inter-

Figure 4: The causal effect of interpersonal uncertainty on allocation decisions



Notes: **Panel A:** The x-axis denotes the amount of money (out of $10) allocated to the group with a lower variation in willingness-to-work (WTW). In **Panels B to D**, the x-axis denotes the amount allocated to the group whose members share a social group with the decision-maker instead of the group whose members are from a different social group. In Panel B, all group members have the same WTW. In Panel C, the outgroup has the higher WTW variation. In Panel D the ingroup has the higher variation. The red line denotes the even split, the blue line the average allocation. In all panels, the binwidth is 10. Decisions are pooled across two social groups (shared hobbies/interests and political views), displaying $n = 120$ (Panel A) and $n = 240$ (Panel B to D) decisions made by 120 subjects.

personal uncertainty. However, we find no evidence for such a confound, as the following test shows. We define *WTW-distance* (i.e., dissimilarity) as the absolute difference between the DM's own WTW and Group A's WTW (12). Regressing the amount given to Group A on this *WTW-distance* reveals a weak and statistically insignificant relationship ($\beta = 0.017$, $p = 0.50$). Using alternative definitions of *WTW-distance* based on Group B WTWs (4 or 22) also yields non-significant results ($\beta = 0.02$, $p = 0.19$ and $\beta = -0.05$, $p = 0.16$, respectively). We therefore conclude that subjects do not appear to base their choices on the WTW-similarity or dissimilarity of the recipients.

3.2.3 The quantitative importance of interpersonal uncertainty

Next, using the data from the previous section, we quantify the relative importance of interpersonal uncertainty and ingroup preferences in driving ingroup favoritism, using a reduced-form analysis, a type analysis, and a structural model of prosocial decision-making.

Reduced form analysis. To estimate the effects of uncertainty and preferences on ingroup preferences, we use the following model: $alloc_{i,d} = c_0 + c_1 unc_d + \varepsilon_{i,d}$, in which $alloc_{i,d}$ denotes the allocation to the ingroup by individual i in decision d. We normalize the variable by subtracting 5 from the actual giving so that the 50:50 split benchmark implies $alloc_{id} = 0$.

The variable unc_d is equal to 1 for decisions with *Higher WTW uncertainty on outgroup members*, equal to -1 for decisions with *Higher WTW uncertainty on ingroup members*, and equal to 0 when uncertainty is absent for both groups. Thus, c_1 measures how much the allocation is affected by uncertainty in either direction. Accordingly, ingroup preferences are measured by the constant c_0 that captures how much more subjects allocate to the ingroup on average in the absence of interpersonal uncertainty ($unc_d = 0$).

We find that both ingroup preferences and interpersonal uncertainty significantly influence behavior. Pooled across groups, the coefficient on preferences is $c_0 = 0.666$, and the coefficient on uncertainty is $c_1 = 0.742$, both being significant at the 0.001 level (see Appendix Table A.1).

When we run the regression separately based on the group identity, we find that the influence of ingroup preferences on behavior is stronger for political views ($c_0^{Pol} = 0.964$) than hobbies/interests ($c_0^{Hobb} = 0.367$), while the influence of uncertainty is similar in both cases ($c_1^{Pol} = 0.720$ and $c_1^{Hobb} = 0.764$, respectively). These results demonstrate that, depending on the social group studied, the influence of uncertainty on behavior can be stronger or weaker than the influence of preferences. Similarly, while we fix the extent of interpersonal uncertainty in this treatment, the extent and hence influence of interpersonal uncertainty would naturally vary across decisions.

Subject-level type analysis. We further exploit the within-subject structure to identify distinct behavioral types. We say a subject reveals a *group-based preference* if they choose a different allocation than 50:50 in *at least one* of the two decisions that provided *group information without uncertainty*. We say that a subject *responds to interpersonal uncertainty* if, for both social groups (interest and political views), their *outgroup uncertainty* decision was different from their *ingroup uncertainty* decision.[27] With this categorization, we find that 20% of subjects neither respond to uncertainty nor display a group preference. 33% of subjects respond to uncertainty but do not display a group preference, while 17% do not respond to uncertainty but display a group preference. Finally, 31% both respond to uncertainty and display a preference. Hence, interpersonal uncertainty is relevant for 64% of all subjects, while group preferences are relevant for 48%.

A complementary approach to the previous reduced form analysis is the use of a structural model, explained next.

Structural model setup. Suppose the representative DM has to distribute M units between two individuals 1 and 2, whose group identity can take one of three values: $G(i) \in \{in, out, \emptyset\}$, where $\emptyset$ means unknown. Suppose the DM believes that the valuations of money

[27] Thus, we use a more conservative identification criterion for the response to uncertainty, because we require subjects to respond to uncertainty across both pairs of choices. In contrast, for identifying group-based preferences only one choice needs to be different from the "no favoritism" benchmark. In total, 79% of subjects respond to at least one change in uncertainty from ingroup uncertainty to outgroup uncertainty.

received by the two individuals 1 and 2 are distributed as f_1 and f_2 respectively, and suppose $x_{IU}(f_1, f_2, \gamma)$ is the choice that maximizes the expected (utilitarian) utility (see equation 1) given the two distributions. To parameterize risk aversion, we assume CRRA utility $u(w) = \frac{w^{1-\gamma}}{1-\gamma}$, with γ as the risk aversion parameter. In our experiment, under the assumption that the valuation of money is measured as WTW, f_1 and f_2 is either the degenerate lottery $L_1 = (12, 1)$ (tasks) or the 50-50 lottery $L_2 = (h = 22, 0.5; l = 4, 0.5)$. Under these distributions and under CRRA risk-preferences, x_{IU} has the following closed form solution:

$$x_{IU}\left(f_1 = (m, 1), f_2 = (h, 0.5; l, 0.5), \gamma\right) = \frac{M \cdot h \left(\frac{m-l}{h-m}\right)^{\frac{1}{\gamma}} - M \cdot l}{(m-l) + (h-m)\left(\frac{m-l}{h-m}\right)^{\frac{1}{\gamma}}}$$

Finally, we assume that the DM's optimal allocation to individual 1 in observation j is as follows:

$$x_{1j} = \begin{cases} \frac{M}{2} + b + \varepsilon_j & \text{if } f_1 = f_2, G(1) = in, G(2) = out \\ x_{IU}\left(f_1, f_2, \gamma\right) + \varepsilon_j & \text{if } f_1 \neq f_2, G(1) = G(2) \\ a_{IU} \cdot x_{IU}\left(f_1, f_2, \gamma\right) + a_{ING} \cdot \left(\frac{M}{2} + b\right) + \varepsilon_j & \text{if } f_i \neq f_o, G(1) = in, G(2) = out \end{cases}$$

In the first case of symmetric interpersonal uncertainty, the ingroup preference factor b alone determines the allocation. The normal noise parameter $\varepsilon_j \sim N(0, \sigma^2)$ is is i.i.d across observations. In the second case of symmetric group information, interpersonal uncertainty alone determines the final allocation. When we have a conjunction of the former two cases, the optimal allocation combines the influence of both factors: $a \lesseqgtr 1$ quantifies if the influence of the corresponding channel shrinks ($a < 1$), stays unchanged ($a = 1$), or expands ($a > 1$) when both factors are present.

Structural model results. We jointly estimate $\gamma, b, a_{IU}, a_{ING}, \sigma$ to maximize the likelihood of the observed data. We estimate a CRRA parameter of $\gamma = 0.374$[28] and the extent of pure ingroup preference to be $b = 0.741$. We estimate a weight of $a_{IU} = 0.739$ on interpersonal uncertainty and a weight of $a_{ING} = 0.336$ on ingroup preferences. Thus, when both factors operate simultaneously, the estimated influences of interpersonal uncertainty and ingroup preferences diminish to 74% and 34% of their respective influences when they operated in isolation. For more details on the estimation, see Appendix Table A.2.

This sub-additive feature inherent in the estimates ($a_{IU}, a_{ING} < 1$) helps one interpret our results from the *social* and *non-social* decisions in light of the reduced-form results from the *Ingroup uncertainty* treatment. One might (incorrectly) think that the quantitative similarity between the *social* and *non-social* decisions implies that interpersonal uncertainty is sufficient to explain *all* observed prosocial behavior, which would be at odds not only with the previous literature but also with our results in the *Group information without uncertainty*

[28]This implies a higher risk-aversion towards interpersonal uncertainty than towards monetary risk. For comparison, across 16 studies employing the Gneezy and Potters (1997) investment task over money, which our setup mimics, Crosetto and Filippin (2016) report an average CRRA parameter $\gamma = 0.3$.

decision. However, the sub-additivity feature explains that when one compares a treatment where both factors are present to one where only interpersonal uncertainty matters, the influence of interpersonal uncertainty in the latter treatment expands and thus partly compensates for the lack of group preferences.

4 Self versus other paradigm (Dictator game)

Our experimental design naturally extends to choices involving tradeoffs between one's own utility versus the utility of others (self versus other decisions), as does the idea that interpersonal uncertainty shapes behavior in these tradeoffs.

4.1 Experimental design

Similar to the ingroup versus outgroup case, DMs face a *Self social* and a *Self non-social* decision, in randomized order. Before the *Self non-social* decision, they also complete the valuation task for \$100 Amazon gift card money received 6 weeks later.

***Self social* decision.** For the *Self social* decision, we endow decision-makers with \$100 which they can allocate between themselves and another individual they have been matched with (without any information about group affiliations). The allocated money is paid out to both parties in the form of Amazon gift card money, six weeks from the date of the experiment. Hence, the *Self social* decision is the standard dictator game: it has consequences for the DM as well as the other individual.

***Self non-social* decision.** In the *Self non-social* decisions, choices again have no social consequences, akin to the *Ingroup non-social* decisions. Decision-makers are endowed with $M = \$100$, and are tasked to split it between their own and another individual's valuation, with their payoff depending on the following formula:

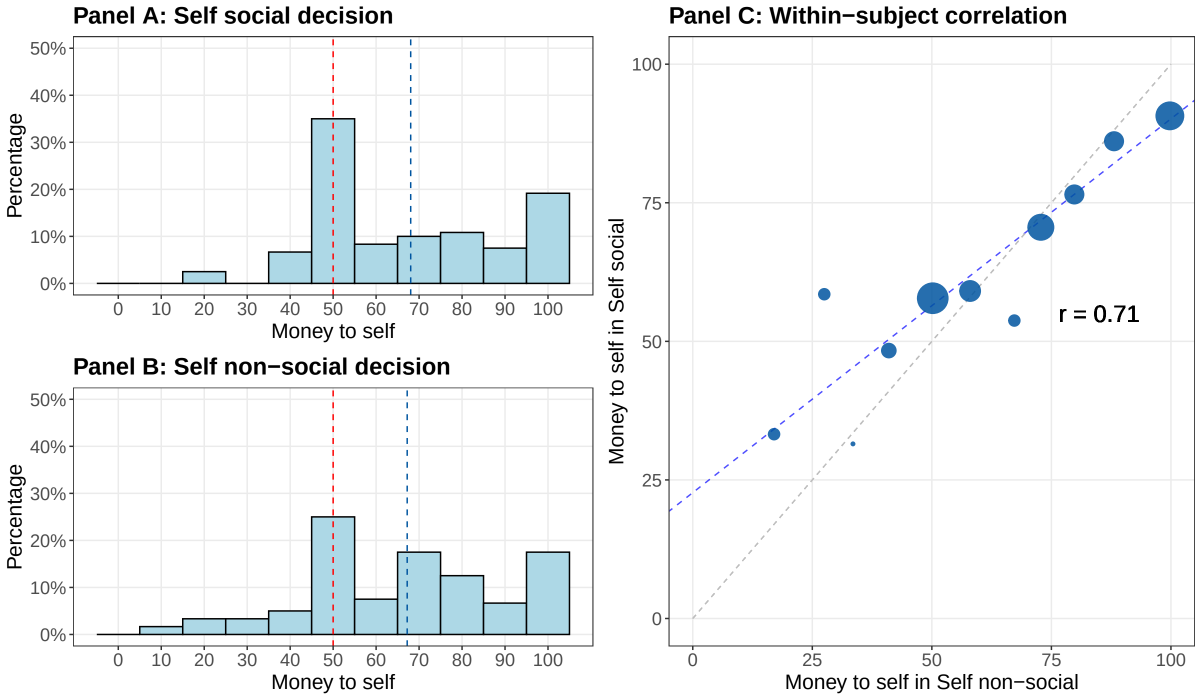
$$\Pi\left(x_{self}, x_{other}\right) = x_{self} \cdot WTP_{self}/100 + x_{other} \cdot WTP_{other}/100$$

where WTP_{self} and WTP_{other} denote the decision-makers and the matched individual's respective WTP for the gift card, and x_{self} and x_{other} are the amounts allocated to the DMs own WTP and to the matched individual's WTP. Accordingly, DMs are incentivized to maximize the sum of their WTP and the WTP of the other individual they are matched with, with both WTPs receiving equal weight. All other elements match the ingroup versus outgroup setting. In total, 120 subjects faced the *Self social* and *Self non-social* decisions.

4.2 Results

***Self social* decision.** In *Self social*, subjects allocate on average \$69.05 to themselves, thereby allocating significantly more to themselves compared to the equal split ($p < 0.001$,

Figure 5: Main results self versus other decision



Notes: **Panel A and B:** Histogram of the *Self social* (Panel A) and *Self non-social* (Panel B) decision. The binwidth is 10. In Panel A, the x-axis denotes the amount of money (out of $100) that subjects allocate to themselves instead of another individual, with decisions having consequences for the ingroup and outgroup members. In Panel B, the x-axis denotes the amount (out of $100) that subjects allocate to their own valuation instead of another individual's valuation, with their payoff depending on the ingroup and outgroup members' valuation of money. The red line denotes the even split, the blue line the average allocation. **Panel C:** Binned scatter plot of the *Self social* and *Self non-social* decision. We divide *Self non-social* choices into 10 bins and plot the average giving to self in *Self social* for each bin. The blue line displays the linear fit of regressing the *Self social* on the *Self non-social* decision. The correlation coefficient is $r = 0.71$. Displayed are $n = 61$ (Panel A), $n = 59$ (Panel B) and $n = 120$ (Panel C) decisions.

one-sample Wilcoxon test). Figure 5 panel A displays the distribution, which replicates the typical pattern of dictator game behavior found in the literature (Engel, 2011). In total, 61% of subjects allocate more money to themselves, 3% allocate more to the other person, and 36% implement the 50/50 split.

***Self non-social* decision.** In *Self non-social*, subjects allocate on average $64.12 to their own WTP, again a significant deviation from the equal split ($p < 0.001$, one-sample Wilcoxon test). As Figure 5 panel B shows, the distribution is also similarly shaped as in the *Self social* case. In total, 58% of subjects allocate more money to their own WTP, 17% allocate more to the other individual's WTP, and 25% implement the 50/50 split.

Comparing *Self social* and *non-social*. Allocations in the *Self non-social* setting closely replicate the behavior we observe in *Self social*. Statistically, we cannot reject that the average amount that subjects allocate to themselves is equal across *social* and *non-social* decisions ($p = 0.27$, unpaired Wilcoxon test). Similarly, we cannot reject that the distribution of allocations is equal across the two decisions ($p = 0.39$, Kolmogorov-Smirnov test). These results also replicate within-subject. Figure 5 panel C binscatter-plots each individual's *social* and *non-social* decision pair. The two decisions are highly correlated at the individual level, with a correlation coefficient of $r = 0.71$ $(p < 0.001)$. Hence, the *Self non-social* decision strongly predicts the *Self social* decision.

Robustness. To show that people are attentive to the *non-social* incentives in the self versus other setting, we use the same incentive treatment as in the ingroup versus outgroup setting (Section 3.1.3) and apply it to the current setting. Our results are similar: subjects understand the incentives and react to them as hypothesized. We provide the results in Appendix Section J.

Result 2. *Self non-social choices replicate the self-favoring behavior found in Self social choices. The distributions are similar and strongly correlated on the individual level.*

5 Measuring perceived interpersonal uncertainty

In this section, we describe how a simple survey question can be used to directly measure perceived interpersonal uncertainty across individuals and decision scenarios. We use this measure to test a central hypothesis of our framework: that subjects perceive higher uncertainty for outgroup members versus ingroup members, and for others versus themselves. Additionally, we demonstrate how this measure predicts prosocial decisions, underscoring its potential as a valuable tool for understanding prosocial behavior.

5.1 Measurement

Elicitation. After the *Ingroup social* (Section 3.1.1) and *Self social* (Section 4.1) decisions, we asked decision-makers the following question, separately about each recipient:

> "How certain are you about how much [the recipient] would value
> [their bonus reward]?"

The question asked about outgroup and ingroup recipients in *Ingroup social* decisions (Section 3.1.1) or self and other recipients in *Self social* decisions (Section 4.1). The bonus reward concerned the object allocated in the previous allocation decision, being Amazon gift card money or cash in the *Gift card* and *Effort* environment, respectively. Subjects responded on an 11-point scale where the 0 and 10 were labeled as *Very uncertain* and *Very certain*, respectively. We recode the variable so that higher values indicate greater uncertainty. We then construct the following individual-level measure of *relative uncertainty*:

$$\begin{aligned}\text{Relative uncertainty} = {} & \text{Uncertainty about outgroup (or other's) valuation} \\ & - \text{Uncertainty about ingroup (or own) valuation}\end{aligned}$$

This method offers a straightforward way to measure subjective interpersonal uncertainty and allows for testing whether subjects perceive systematic differences between recipients.

Results. In *Ingroup social*, a large majority of subjects perceive non-zero uncertainty about the outgroup and ingroup members valuations (in 77% and 70% of cases, respectively). Importantly, subjects perceive higher uncertainty about valuations of outgroup members than ingroup members. Measured in scale points, this difference is 0.76 in the case of hobbies/interests, 1.20 in the case of political views, and 0.73 in the case of religious beliefs. All differences are significantly different from zero ($p < 0.001$, paired Wilcoxon tests). With respect to the distribution of differences at the individual level, in 32% of cases, subjects report higher uncertainty for the outgroup; in 8% they report higher uncertainty for the ingroup and in the remaining 59% cases, subjects report no difference. On the aggregate, the percentages are thus similar to the percentages of behavior in the *Ingroup social* decision, where 44%, 5%, and 51% of subjects show ingroup-favoritism, outgroup-favoritism, and no favoritism, respectively.

Similarly, in *Self social*, 91% of subjects perceive non-zero uncertainty about a randomly selected other person's valuation, and 53% perceive uncertainty about their own valuation. Subjects perceive on average 2.59 scale points higher uncertainty for the other person's valuation than their own, which is again significantly different from the no difference benchmark ($p < 0.001$, paired Wilcoxon tests). In total, 72% of subjects report a higher uncertainty about the other person's valuations, 6% report a higher uncertainty about their own valuation, and 23% report no difference. Recall that in terms of behavior, 61% of subjects allocate more to themselves, 3% more to the other person, and 36% share 50/50.

Predicting choices. Moving to individual-level associations, as previously shown in Figure 1 of the introduction, we find that differences in interpersonal uncertainty strongly predict allocation decisions in both *Ingroup social* and *Self social*. Higher uncertainty about the outgroup's valuation relative to the ingroup's valuation predicts stronger ingroup favoritism ($r = 0.37$, $p < 0.001$), and higher uncertainty about the other individual's valuation relative to own valuation predicts stronger giving to self ($r = 0.35$, $p < 0.001$). Moreover, our measure significantly predicts choices in the *Ingroup non-social* and *Self non-social* ($r = 0.23$, $p < 0.001$ and $r = 0.24$, $p = 0.009$, respectively). For the binned scatter plot, see Appendix Figure B.2.

5.2 Robustness

We designed a robustness treatment, *IU belief measurement* ($n = 120$), to address two potential concerns: (1) the possibility that prior allocation decisions influenced uncertainty elicitation, and (2) the risk that subjects conflated uncertainty with mean beliefs. *IU belief measurement* had no allocation decisions. Instead, subjects first learned how valuations were measured for past participants, then guessed the average valuations of past participants, and afterwards reported their interpersonal uncertainty. This process ensures a clear distinction between average valuations and uncertainty. For example, in the ingroup versus outgroup case, subjects first reported average estimates

> Consider individuals who [share your interests/hobbies / have different interests/hobbies than you]. On average, how much do you think they value [their bonus reward]?

and *then* reported their interpersonal uncertainty about both ingroup and outgroup, using the same format described in Section 5.1.

Results. We start with subjects' reports about uncertainty before moving to estimates about the average. We replicate our uncertainty patterns from Section 5.1. For the ingroup versus outgroup case, subjects again report higher uncertainty about outgroup members for all three social groups (0.81 points higher for hobbies/interests, 0.93 for political views, and 0.66 for religious beliefs, each $p < 0.01$, paired Wilcoxon tests). Importantly, we cannot reject that the distribution of beliefs elicited without prior allocation decision is the same as elicited with prior decision (all $p > 0.10$, Kolmogorov-Smirnov test). Similarly, in the self versus other case, subjects perceive higher uncertainty over the other person's valuation relative to their own (3.11 points higher, $p < 0.001$, paired Wilcoxon test).[29]

Turning to estimates about the average, subjects report ingroup members to have valuations of $87.85 (hobbies/interests), $89.64 (political views), and $89.14 (religious beliefs). For outgroup members, they report valuations of $88.20, $86.69, and $87.85. Thus, on average, subjects' beliefs do not differ between ingroup and outgroup ($p = 0.90$, $p = 0.06$, and $p = 0.37$, paired Wilcoxon tests), and are well calibrated, as the actual average WTP observed in the experiment is $87. The average belief about mean valuations in the self versus other case is $83.43, whereas the actual WTP is $86.69. Thus, on average, subjects believe others to have a slightly lower WTP than themselves ($p = 0.03$, paired Wilcoxon test).

As an additional validation exercise, at the end of *IU belief measurement*, we presented subjects with two artificial groups featuring exogenously induced WTP distributions, where one was a mean-preserving spread of the other. Subjects then reported their perceived interpersonal uncertainty for each group. Overall, 74% (87% respectively) of subjects report strictly (weakly) higher interpersonal uncertainty about the first group, validating the sensitivity of the uncertainty measure. See Appendix L for details.

[29] Regarding the level of uncertainty, in ingroup versus outgroup, in 88% and 85% do people perceive non-zero uncertainty about the outgroup and ingroup, respectively. In the self versus other condition, 98% of participants report non-zero uncertainty regarding a randomly selected other individual's valuation, whereas 74% report such uncertainty regarding their own valuation.

5.3 Changing interpersonal uncertainty

Our results so far demonstrate how self-reported interpersonal uncertainty tracks and predicts prosocial decisions *across* different recipients *within* the same decision situation. Next, we assess whether it also predicts changes in behavior across behavioral interventions, focusing on two prominent interventions: the *identifiable victim effect* and the *contact effect*.

The identifiable victim effect describes the increase in giving when information about a specific, identified individual is provided, compared to general information about a large group (see Lee and Feeley, 2016, for a meta-analysis). The contact effect describes the empirical finding that individuals exhibit greater prosocial behavior towards others after engaging in some form of social interaction (see Clochard, 2024; Lowe, 2025, for recent meta-analyses).

We hypothesize that identifiable victim interventions reduce interpersonal uncertainty by providing concrete details about an individual, making their needs and circumstances easier to imagine with greater confidence. Similarly, social interactions in contact interventions offers firsthand insights into others' life and experiences, directly reducing uncertainty about their preferences and circumstances. This, coupled with our main hypothesis that lower interpersonal uncertainty leads to higher prosocial actions, immediately predicts how these behavioral interventions would shape behavior. In our subsequent experiment, we will test whether self-reported interpersonal uncertainty predicts the effects of these interventions on prosocial behavior; if the data confirm our expectations, it would suggest that part of their effectiveness stems from reducing uncertainty. More broadly, this would point to a new mechanism—interpersonal uncertainty—that helps explain the malleability of prosocial behavior under different interventions.

Design. Each of the two interventions–contact and identifiable victim–features two conditions, a treatment and a control condition. In the two *Contact* conditions, subjects are matched with another individual. In the *Contact control* condition, subjects then play a dictator game, allocating $100 between themselves and their matched individual. In *Contact treatment*, subjects interact with their matched individual prior to allocating the money. Specifically, subjects are given five ice-breaker-type questions (Example: "What's a memorable experience that has shaped who you are today?"). They choose three questions to answer, and their responses are shared with their matched individual. In turn, the matched individual's answers to the same questions are shown to the subject. This procedure simulates contact through a conversational exchange.

In the two identifiable victim (IV) conditions, subjects allocate $100 between themselves and recipients living in Malawi through a donation to the charity GiveDirectly. In both the control and treatment conditions, subjects first receive a general introduction about the charity and Malawi. In the *IV control* condition, they then receive some general information about the recipients in Malawi. In *IV treatment*, they receive similar informational content but through the first-hand account from a specific, identified Malawi recipient who shared their story on the GiveDirectly website.

In each condition, we elicit subjects' interpersonal uncertainty after their allocation decision. Specifically, in the contact condition we elicit interpersonal uncertainty about their matched partner and in the identifiable victim conditions about the recipients in Malawi. In total, 602 subjects participated in the experiment, 149 each in *Contact treatment* and *Contact control*, and 152 each in *IV treatment* and *IV control*.

Results. First, we replicate the effectiveness of the interventions on increasing prosocial behavior. While subjects in *Contact control* allocate \$69.04 to themselves, in *Contact treatment* giving to self decreases to \$53.48 ($p < 0.001$, two-sample Wilcoxon test). In *IV control*, subjects allocate \$56.70 to themselves, which decreases to \$48.74 in *IV treatment* ($p = 0.046$, two-sample Wilcoxon test).

Importantly, we find that both interventions also significantly influence perceived interpersonal uncertainty. Average interpersonal uncertainty about the matched partner in contact control is 3.22 points, which decreases to 2.68 in contact treatment ($p = 0.034$, two-sample Wilcoxon test). Similarly, while average uncertainty about the recipients in *IV control* is 3.14, it decreases to 1.97 in *IV treatment* ($p < 0.001$, two-sample Wilcoxon test). Moreover, our measure of interpersonal uncertainty predicts the extent of self-favoritism in both instances, as displayed in Appendix Figure B.3 (identifiable victim: $r = 0.37$, $p < 0.001$, contact: $r = 0.33$, $p < 0.001$).

In summary, both interventions not only increase giving but also significantly decrease perceived uncertainty, providing insight into how these interventions may promote prosocial behavior by reducing interpersonal uncertainty.

6 Discussion

Extension to fairness behavior. The focus of our framework and experiments has been on settings where decision-makers (DMs) face *differential interpersonal uncertainty* across recipients. Specifically, cases where the perceived valuation distribution of one recipient is a mean-preserving spread of another. In some cases, however, DMs might perceive equal uncertainty but different means about two recipients, i.e., a *mean shift* in the perceived valuation distributions. We explore such a case in Appendix Section K, where we apply our framework to study *merit-based fairness*.

There, we examine a setting in which the endowment to be allocated in a dictator game is either *earned* or randomly assigned (*windfall*). Prior literature consistently finds that dictators allocate more generously when the recipient has earned the endowment.[30] Interper-

[30] Ruffle (1998), Cherry (2001), Cherry, Frykblom, and Shogren (2002), Cherry and Shogren (2008), Oxoby and Spraggon (2008), and Krupka and Weber (2013)

sonal uncertainty offers a belief-based explanation for this behavior: if DMs believe that, on average, a recipient experiences greater *disutility from losing earned income* than *utility from gaining unearned income* (i.e., a gain–loss asymmetry), they will perceive the valuation distribution of the "earner" as *shifted to the right*—thereby justifying more generous allocations. Using a separate experiment, we provide empirical support for this channel.

Interactions of subjective with objective uncertainty. A key feature of interpersonal uncertainty is its subjective nature: one can perceive it even when there is no objective uncertainty about the outcomes of a prosocial action. It turns out that when objective uncertainty is introduced, it can interact in interesting ways with interpersonal uncertainty. One such interaction can be observed in decision situations where DMs appear to use objective uncertainty to justify acting selfishly – a pattern known as *excuse-driven selfish behavior* (Bénabou and Henkel, 2025). For example, Exley (2016) finds that individuals are more tolerant of risk when it affects their own payoffs than when it affects the payoffs of others. As a result, they are less willing to give when recipients face uncertainty. This asymmetric risk sensitivity is interpreted as evidence that, driven by concerns about self-image, people use uncertainty as an "excuse" to behave selfishly.

Interpersonal uncertainty offers a complementary explanation for why individuals might be more tolerant of risk when it affects their own payoffs than when it affects the payoffs of others. We relegate the formal proof to Appendix Section M, and sketch an intuitive explanation here. To illustrate, we display in Table 1 the four decision types (labeled D1-D4) from Exley (2016). Each decision type was elicited as an indifference point between the left and the right options, on a multiple price list. The decisions varied in whether the payoffs are directed toward oneself or others, and whether the uncertainty involved is objective (OU), interpersonal (IU), or both.

D3 and D4 capture similar tradeoffs: D3 features only OU (no IU). D4 features IU on both the certain and uncertain option, effectively "canceling out" IU from the tradeoff. As a result, D3 and D4 both isolate identical tradeoffs driven only by OU, and indeed, subjects in Exley (2016) exhibit identical choice patterns across D3 and D4. In contrast, D1 and D2 capture very different tradeoffs. In D1, the IU and OU both apply to the uncertain outcome (benefiting the other), compounding uncertainty. As a result, risk-averse individuals are willing to settle for a smaller, certain outcome (benefiting the self) – thereby reinforcing behavior that appears selfish. In D2, IU is instead present on the certain outcome (benefiting the other), and OU is on the uncertain outcome (benefiting the self). IU here undermines the perceived "certainty" of the other-benefiting option, making it less appealing. As a result, DMs demand a larger certain (other-benefiting) payoff to reach indifference, making them appear more generous in D2 than in D1. Thus, IU can explain the asymmetric behavior documented in Exley (2016).

Note that this novel explanation does not rule out existing explanations, for example, excuse-driven selfish behavior – rather, it suggests that interpersonal uncertainty and excuse-driven motivations may jointly influence behavior. Indeed, Exley (2016) additionally shows that the asymmetric risk response pattern is stronger for subjects that prefer to avoid payoff-relevant information in the moral wiggle room game (Dana, Weber, and Kuang, 2007), providing direct evidence for excuse-driven selfish behavior.

Table 1: Interactions of objective and subjective uncertainty

Decision	**Tradeoff**		**Intuition of tradeoff**
	Certain outcome	**Uncertain outcome**	
D1	Own pay	Other's pay (OU+IU)	IU *strengthens* OU
D2	Other's pay (IU)	Own pay (OU)	IU *weakens* OU
D3	Own pay	Own pay (OU)	No IU on either option
D4	Other's pay (IU)	Other's pay (OU+IU)	IU cancels from both options

Notes: Tradeoffs in the four decision types in Exley (2016). IU = Interpersonal uncertainty; OU = Objective uncertainty. D3 and D4 are symmetric (IU present on both or neither side). D1 and D2 are asymmetric: in D1, IU and OU reinforce selfishness, whereas in D2 they pull in opposite directions.

A discussion on motivated beliefs. One might also wonder whether differences in self-reported interpersonal uncertainty are driven entirely by *motivated reasoning* – that is, belief distortions that serve one's preferences, such as ingroup favoritism. Under this theory, individuals should systematically manipulate their beliefs about the outgroup's utility – both the mean and the uncertainty – whenever doing so favors their preferred outcome. Moreover, such manipulation should increase with the potential gains from it.

This framework yields three testable predictions. First, subjects making *Ingroup social* allocation decisions should *inflate* their perceived uncertainty about the outgroup's valuation more than those in the *IU belief measurement* condition where no allocation decision is made. Second, motivated subjects should also deflate the mean of their outgroup beliefs in ways that justify giving more to the ingroup. Third, if elevated outgroup uncertainty is driven entirely by motivated belief distortion, it should *vanish* in the *Ingroup non-social* condition. This is because subjects have a direct incentive to hold correct beliefs, as doing so maximizes monetary payoffs in the *Non-social* decision.[31]

However, our data rejects all three predictions. First, we find no significant differences between beliefs elicited with or without prior allocation choice (Section 5.2). Second, we find no significant differences in mean beliefs between ingroup and outgroup (Section 5.2). Third, we find behavior consistent with perceived differences in interpersonal uncertainty

[31]Indeed, it has been shown that incentives for belief accuracy mitigate the role of motivated reasoning (e.g., Zimmermann, 2020; Sprengholz et al., 2023).

in *Ingroup non-social* (Section 3.1.2), which, in turn, is significantly predicted by subjects' stated beliefs about the differences (Section 5.1).

Perhaps most importantly, motivated reasoning or excuse-driven behavior should also give rise to information avoidance. That is, decision-makers who intend to favor their own group and hold motivated beliefs may prefer to avoid information about recipients' utility that could challenge those beliefs. However, we find that a majority of subjects (81%) prefer to receive information about recipients' WTW – which reduces interpersonal uncertainty – rather than information about which recipient is part of their ingroup and which is part of their outgroup (see Appendix I).

7 Conclusion

In this paper, we provide a conceptual framework and implement a series of experiments documenting how interpersonal uncertainty bolsters ingroup favoritism, weakens altruistic giving, and shapes redistributive behavior. We show that a significant degree of heterogeneity in prosocial behavior, both within a given decision setting and between different settings, is driven by people's differential response to interpersonal uncertainty. As a consequence, precise identification of social preferences from prosocial behavior requires explicit accounting for interpersonal uncertainty. Otherwise, depending on the nature of interpersonal uncertainty, parameters of social preferences may be over- or underestimated. We also demonstrate an experimental design to disentangle uncertainty from preferences: a researcher can exogenously vary interpersonal uncertainty to explicitly measure and control for it. For instance, in our experiment, we provide subjects with information so that interpersonal uncertainty switches between recipients, or is balanced among recipients.

Finally, our framework supports the idea that prosocial behavior is malleable. Under the assumption of "exposure reduces interpersonal uncertainty", it helps explain the dynamics of prosocial behavior in response to intergroup contact created by spatial proximity (Bursztyn et al., 2024), shared classes (Rao, 2019), shared living (Corno, La Ferrara, and Burns, 2022), sports events (Mousa, 2020; Lowe, 2021) and attending youth camps (Ghosh et al., 2024). Indeed, we provide evidence that a stylized contact intervention not only increases prosocial behavior, but also decreases interpersonal uncertainty. Similarly, our conceptual framework and results vindicate how people's degree of favoritism towards specific groups varies based on their closeness (Fong and Luttmer, 2009), salience of shared experiences (McLeish and Oxoby, 2011), or (perceived) similarity (Goeree et al., 2010) to ingroup members. Overall, our results suggest that targeting and reducing interpersonal uncertainty could foster prosocial behavior, bridge animosities, and decrease intergroup conflict.

References

Alma, Ingvild, Alexander W Cappelen, and Bertil Tungodden (2020). "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy* 128 (5): 1753–88. [59]

Amador, Manuel, Ivan Werning, and George-Marios Angeletos (2006). "Commitment vs. Flexibility." *Econometrica* 74 (2): 365–96. [7]

Andries, Marianne, Leonardo Bursztyn, Thomas Chaney, Milena Djourelova, and Alex Imas (2024). "In their shoes: Empathy through information." *Working Paper*, [7]

Bénabou, Roland, Armin Falk, and Luca Henkel (2024). "Ends versus Means: Kantians, Utilitarians and Moral Decisions." *Working Paper*, [59]

Bénabou, Roland, and Luca Henkel (2025). "Identity as Self-Image." *Working Paper*, [33]

Berg, Joyce, John W. Dickhaut, and Kevin McCabe (1995). "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10 (1): 122–42. [7]

Böhm, Robert, Hannes Rusch, and Jonathan Baron (2020). "The Psychology of Intergroup Conflict: A Review of Theories and Measures." *Journal of Economic Behavior and Organization* 178: 947–62. [6]

Brock, J. Michelle, Andreas Lange, and Erkut Y Ozbay (2013). "Dictating the Risk: Experimental Evidence on Giving in Risky Environments." *American Economic Review* 103 (1): 415–37. [7]

Bursztyn, Leonardo, Thomas Chaney, Tarek A. Hassan, and Aakaash Rao (2024). "The Immigrant Next Door." *American Economic Review* 114 (2): 348–84. [35]

Camerer, Colin F. (1995). "Individual Decision Making." In *Handbook of Experimental Economics*. John H. Kagel and Alvin E. Roth, ed. vol. 2, Princeton, NJ: Princeton University Press, 587–703. [58]

Cappelen, Alexander W, Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden (2022). "A Meritocratic Origin of Egalitarian Behaviour." *Economic Journal* 132 (646): 2101–17. [7]

Cappelen, Alexander W., Thomas De Haan, and Bertil Tungodden (2024). "Fairness and Limited Information: Are People Bayesian Meritocrats?" *Journal of Public Economics* 233 (5): 105097. [7]

Cappelen, Alexander W., Ranveig Falch, and Bertil Tungodden (2020). "Fair and Unfair Income Inequality." In *Handbook of Labor, Human Resources and Population Economics*. Klaus F. Zimmermann, ed. Cham: Springer International Publishing, 1–25. [57]

Capraro, Valerio, Joseph Y. Halpern, and Matjaz Perc (2024). "From Outcome-Based to Language-Based Preferences." *Journal of Economic Literature* 62 (1): 115–54. [7, 11]

Chakraborty, Anujit (2021). "Present bias." *Econometrica* 89 (4): 1921–61. [7, 9]

Chakraborty, Anujit, Yoram Halevy, and Kota Saito (2020). "The relation between behavior under risk and over time." *American Economic Review: Insights* 2 (1): 1–16. [7]

Charness, Gary, and Yan Chen (2020). "Social Identity, Group Behavior, and Teams." *Annual Review of Economics* 12: 691–713. [6]

Chen, Daniel L., Martin Schonger, and Chris Wickens (2016). “oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments.” *Journal of Behavioral and Experimental Finance* 9: 88–97. [14]

Chen, Yan, and Sherry Xin Li (2009). “Group Identity and Social Preferences.” *American Economic Review* 99 (1): 431–57. [11]

Chen, Yiting, and Songfa Zhong (2025). “People Are More Moral in Uncertain Environments.” *Econometrica* forthcoming: [7]

Cherry, Todd L, Peter Frykblom, and Jason F Shogren (2002). “Hardnose the Dictator.” *American Economic Review* 92 (4): 1218–21. [32]

Cherry, Todd L. (2001). “Mental Accounting and Other-Regarding Behavior: Evidence from the Lab.” *Journal of Economic Psychology* 22 (5): 605–15. [32]

Cherry, Todd L., and Jason F. Shogren (2008). “Self-Interest, Sympathy and the Origin of Endowments.” *Economics Letters* 101 (1): 69–72. [32]

Clochard, Gwen-Jiro (2024). “Reducing prejudice: A Meta-Analysis and Experiment on the Contact Hypothesis.” *Working Paper*, [31]

Corno, Lucia, Eliana La Ferrara, and Justine Burns (2022). “Interaction, Stereotypes, and Performance: Evidence from South Africa.” *American Economic Review* 112 (12): 3848–75. [35]

Crosetto, Paolo, and Antonio Filippin (2016). “A theoretical and experimental appraisal of four risk elicitation methods.” *Experimental Economics* 19: 613–41. [25]

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2007). “Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness.” *Economic Theory* 33 (1): 67–80. [34]

Dasgupta, Partha, and Eric Maskin (2005). “Uncertainty and hyperbolic discounting.” *American Economic Review* 95 (4): 1290–99. [7]

Engel, Christoph (2011). “Dictator Games: A Meta Study.” *Experimental Economics* 14 (4): 583–610. [27]

Enke, Benjamin (2023). “Moral Boundaries.” *Annual Review of Economics* forthcoming: [6]

Enke, Benjamin, and Thomas Graeber (2023). “Cognitive Uncertainty.” *Quarterly Journal of Economics* 138 (4): 2021–67. [7]

Enke, Benjamin, Thomas Graeber, and Ryan Oprea (2024). “Complexity and Time.” *Journal of the European Economic Association* forthcoming: [12]

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann (2022). “Moral Universalism: Measurement and Economic Relevance.” *Management Science* 68 (5): 3590–603. [11, 12, 15]

Exley, Christine L. (2016). “Excusing Selfishness in Charitable Giving: The Role of Risk.” *Review of Economic Studies* 83 (2): 587–628. [7, 33, 34, 62]

Fehr, Ernst, and Gary Charness (2025). “Social Preferences: Fundamental Characteristics and Economic Consequences.” *Journal of Economic Literature* 63 (2): 440–514. [7]

Fehr, Ernst, and Simon Gächter (2000). "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives* 14 (3): 159–81. [7]

Fong, Christina M, and Erzo F. P Luttmer (2009). "What Determines Giving to Hurricane Katrina Victims? Experimental Evidence on Racial Group Loyalty." *American Economic Journal: Applied Economics* 1 (2): 64–87. [35]

Gabaix, Xavier, and David Laibson (2022). "Myopia and Discounting." *Working Paper*, [9]

Ghosh, Arkadev, Prerna Kundu, Matt Lowe, and Gareth Nellis (2024). "Creating Cohesive Communities: A Youth Camp Experiment in India." *Working Paper*, [35]

Gill, David, and Victoria Prowse (2012). "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review* 102 (1): 469–503. [13]

Gill, David, and Victoria Prowse (2019). "Measuring Costly Effort Using the Slider Task." *Journal of Behavioral and Experimental Finance* 21 (3): 1–9. [13, 21]

Gneezy, Uri, and Jan Potters (1997). "An experiment on risk taking and evaluation periods." *quarterly journal of economics* 112 (2): 631–45. [25]

Goeree, Jacob K, Margaret A McConnell, Tiffany Mitchell, Tracey Tromp, and Leeat Yariv (2010). "The 1/ d Law of Giving." *American Economic Journal: Microeconomics* 2 (1): 183–203. [35]

Halevy, Yoram (2008). "Strotz meets Allais: Diminishing impatience and the certainty effect." *American Economic Review* 98 (3): 1145–62. [7]

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood (2019). "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22 (1): 129–46. [6]

Kappes, Andreas, Anne-Marie Nussberger, Nadira S. Faber, Guy Kahane, Julian Savulescu, and Molly J. Crockett (2018). "Uncertainty about the Impact of Social Decisions Increases Prosocial Behaviour." *Nature Human Behaviour* 2 (8): 573–80. [7]

Krupka, Erin L., and Roberto A. Weber (2013). "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11 (3): 495–524. [32, 57]

Lee, Seyoung, and Thomas Hugh Feeley (2016). "The Identifiable Victim Effect: A Meta-Analytic Review." *Social Influence* 11 (3): 199–215. [31]

Lowe, Matt (2021). "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *American Economic Review* 111 (6): 1807–44. [35]

Lowe, Matt (2025). "Has Intergroup Contact Delivered?" *Working Paper*, [31]

McLeish, Kendra N., and Robert J. Oxoby (2011). "Social Interactions and the Salience of Social Identity." *Journal of Economic Psychology* 32 (1): 172–78. [35]

Mousa, Salma (2020). "Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq." *Science* 369 (6505): 866–70. [35]

Oprea, Ryan (2024). "Decisions under Risk Are Decisions under Complexity." *American Economic Review* 114 (12): 3789–811. [12]

Oxoby, Robert J., and John Spraggon (2008). “Mine and Yours: Property Rights in Dictator Games.” *Journal of Economic Behavior & Organization* 65 (3-4): 703–13. [32, 57]
Rao, Gautam (2019). “Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools.” *American Economic Review* 109 (3): 774–809. [35]
Ruffle, Bradley J. (1998). “More Is Better, But Fair Is Fair: Tipping in Dictator and Ultimatum Games.” *Games and Economic Behavior* 23 (2): 247–65. [32]
Shayo, Moses (2020). “Social Identity and Economic Policy.” *Annual Review of Economics* 12: 355–89. [6]
Sozou, Peter D. (1998). “On hyperbolic discounting and uncertain hazard rates.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265 (1409): 2015–20. [7]
Sprengholz, Philipp, Luca Henkel, Robert Böhm, and Cornelia Betsch (2023). “Historical Narratives about the COVID-19 Pandemic Are Motivationally Biased.” *Nature* 623 (7987): 588–93. [34]
Tabellini, Guido (2008). “The Scope of Cooperation: Values and Incentives.” *Quarterly Journal of Economics* 123 (3): 905–50. [11]
Tungodden, Bertil, and Alexander W. Cappelen, eds. (2019). *The Economics of Fairness*. Edward Elgar Publishing. [57]
Umer, Hamza (2023). “Effectiveness of Random Payment in Experiments: A Meta-Analysis of Dictator Games.” *Journal of Economic Psychology* 96 (6): 102608. [14]
Zimmermann, Florian (2020). “The Dynamics of Motivated Beliefs.” *American Economic Review* 110 (2): 337–63. [34]

Appendix: Proofs

Proof of Theorem 1. For simplicity, whenever possible we will write $E_{v_i \sim f_i}$ simply as E_{f_i}. From the utility expression, we get

$$\begin{aligned} \frac{d}{dx}EU(x) &= E_{f_1,f_2}\left((v_1 - v_2)U'\left(v_1 x + v_2(100-x)\right)\right) \\ \frac{d^2}{dx^2}EU(x) &= E_{f_1,f_2}\left((v_1 - v_2)^2 U''\left(v_1 x + v_2(100-x)\right)\right) < 0 \end{aligned}$$

as $U'' < 0$ and $f_1, f_2 \geq 0$. $\frac{d^2}{dx^2}EU(x)$ being strictly positive implies that $\frac{d}{dx}EU(x) = 0$ must be obtained at a unique point. Evaluating the first derivative at $x = 50$, we get

$$\frac{d}{dx}EU(x)|_{x=50} = E_{f_1,f_2}(v_1 - v_2)U'\left(50v_1 + 50v_2\right) \tag{2}$$

(i) When f_1 and f_2 are identical, then we can also rewrite

$$\begin{aligned} \frac{d}{dx}EU(x) &= E_{v_1\sim f_2}E_{v_2\sim f_1}(v_1 - v_2)U'\left(v_1 x + v_2(100-x)\right) \\ &= E_{v_2\sim f_2}E_{v_1\sim f_1}(v_2 - v_1)U'\left(v_2 x + v_1(100-x)\right) \\ &= E_{f_1,f_2}(v_2 - v_1)U'\left(v_2 x + v_1(100-x)\right) \end{aligned}$$

where the first step integrates v_1 over f_2 and v_2 over f_1 instead, the second step interchanges the names of variables (v_1 and v_2) of integration. The last step interchanges the order of integration. Evaluating the final expression at $x = 50$, we get

$$\frac{d}{dx}EU(x)|_{x=50} = E_{f_1,f_2}(v_2 - v_1)U'\left(50v_1 + 50v_2\right) \tag{3}$$

Equations 2 and 3 together imply $\frac{d}{dx}EU(x)|_{x=50} = -\frac{d}{dx}EU(x)|_{x=50} = 0$.

(ii) When f_2 is a mean-preserving spread of f_1, then there exists a random variable $z \sim f_z$ with zero expectation conditional on any given value of v_1, such that v_2 has the same distribution as $v_1 + z$, or in other words, $v_2 =^d v_1 + z$. Therefore, we can replace v_2 by a variable $w_1 + z$ where w_1 and v_1 both have identical distribution f_1.

$$\frac{d}{dx}EU(x) = E_{v_1\sim f_1}E_{w_1\sim f_1}E_{z|v_1,w_1}(v_1 - w_1 - z)U'\left(v_1x + (w_1 + z)(100 - x)\right) \tag{4}$$

$$\frac{d}{dx}EU(x)|_{x=50} = E_{f_1}E_{f_1}E_{z|v_1,w_1}(v_1 - w_1 - z)U'\left(50v_1 + 50w_1 + 50z\right) \tag{5}$$

Because $E_{f_1}E_{f_1}$ is integrating with respect to two identical independent distributions, we can interchange their variable names (w_1 and v_1) in Equation 5:

$$\frac{d}{dx}EU(x)|_{x=50} = E_{f_1,f_1}E_{z|v_1,w_1}(w_1 - v_1 - z)U'\left(50v_1 + 50w_1 + 50z\right) \tag{6}$$

Adding equations 5 and 6, and then using law of iterated expectations:

$$\begin{aligned}
2\frac{d}{dx}EU(x)|_{x=50} &= -E_{f_1,f_1}\left(E_{z|v_1,w_1}2zU'\left(50v_1 + 50w_1 + 50z\right)\right) \\
&> -E_{f_1,f_1}\left(E_{z|v_1,w_1}2zU'\left(50v_1 + 50w_1\right)\right) \\
&= -E_{f_1,f_1}U'\left(50v_1 + 50w_1\right)\left(E_{z|v_1,w_1}2zf_z(z|v_1)dz\right) \\
&= 0
\end{aligned}$$

The inequality uses the fact: $zU'\left(50v_1 + 50w_1 + 50z\right) < zU'\left(50v_1 + 50w_1\right)$ for both $z > 0$ and $z < 0$. The last step follows from the fact that $E_{z|v_1,w_1}z = 0$. Therefore, $\frac{d}{dx}EU(x)|_{x=50} > 0$, and thus, the optimal allocation $x^* > 50$. Next,

$$\begin{aligned}
\frac{d}{dx}EU(x)|_{x=100} &= E_{f_1}E_{f_1}E_{z|v_1,w_1}(v_1 - w_1 - z)U'\left(100v_1\right) \\
&= E_{f_1}E_{f_1}U'\left(100v_1\right)E_{z|v_1,w_1}(v_1 - w_1 - z) \\
&= E_{f_1}E_{f_1}U'\left(100v_1\right)\left(v_1 - w_1\right) \\
&= E_{f_1}E_{f_1}U'\left(100w_1\right)\left(w_1 - v_1\right) \\
&= \frac{1}{2}E_{f_1}E_{f_1}[U'\left(100w_1\right)\left(w_1 - v_1\right) + U'\left(100v_1\right)\left(v_1 - w_1\right)] \\
&= \frac{1}{2}E_{f_1}E_{f_1}(U'(100w_1) - U'(100v_1))(w_1 - v_1) \\
&< 0
\end{aligned}$$

Step 1 replaces $x = 100$ into the expression of $\frac{d}{dx}EU(x)$ derived at the beginning of the proof. Step 2 uses that $U'\left(100v_1\right)$ is independent of z. Step 3 uses $E_{z|v_1,w_1}z = 0$. Step 4 uses the property that v_1, w_1 are drawn i.i.d from f_1, and hence those two variable names can be interchanged. Step 5 uses the average of the two expressions from the previous lines. The last step uses the property that U' is decreasing.

As $\frac{d}{dx}EU(x)|_{x=100} < 0$, the concavity of the expression implies that $\frac{d}{dx}EU(x) = 0$ must be obtained at some $50 < x < 100$.

(iii) The first derivative of the objective function, evaluated at x^*, should be zero.

$$E_{f_1,f_2}(v_1 - v_2 - c)U'(x^*v_1 + (100 - x^*)(v_2 + c)) = 0 \tag{7}$$

First, taking the implicit derivative of the last equation w.r.t c, and then re-arranging:

$$E_{f_1,f_2}[-U' + (v_1 - v_2 - c)^2 \frac{dx^*}{dc} U'' + (v_1 - v_2 - c)(100 - x^*)U''] = 0$$

Next, we re-arrange and then bound $\frac{dx^*}{dc}$ in 6 steps as explained below. Under CARA,

$$\begin{aligned}
\frac{dx^*}{dc} &= \frac{E_{f_1,f_2} - U'}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} + \frac{E_{f_1,f_2}(v_1 - v_2 - c)(100 - x^*)U''}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} \\
&= \frac{E_{f_1,f_2} - U'}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} + \frac{E_{f_1,f_2}(v_1 - v_2 - c)(100 - x^*) \times \frac{U''_{100}}{U'_{100}} U'}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} \\
&= \frac{E_{f_1,f_2} - U'}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} + \frac{(U''_{100})(100 - x^*)}{U'_{100}} \frac{E_{f_1,f_2}(v_1 - v_2 - c)U'}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} \\
&= \frac{E_{f_1,f_2} - U'}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} + \frac{(U''_{100})(100 - x^*)}{U'_{100}} \times \frac{0}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} < 0
\end{aligned}$$

The second step utilizes the assumption of constant absolute risk aversion: $\frac{U''}{U'} = \frac{U''_{100}}{U'_{100}}$, and hence, $U'' = \frac{U''_{100}}{U'_{100}} U'$. The third step simply reorganizes the numerator in the second additive term. The fourth step uses equation 7 to set $E_{f_1,f_2}(v_1 - v_2 - c)U'$ to zero. The last step uses $U' > 0, U'' < 0$.

Under CRRA preferences,

$$\begin{aligned}
\frac{dx^*}{dc} &= \frac{E_{f_1,f_2}[-U' + (v_1 - v_2 - c)(100 - x^*)U'']}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} \\
&= \frac{E_{f_1,f_2}[-U' - (x^*v_1 + (100 - x)(v_2 + c))U'' + 100v_1U'']}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} \\
&= \frac{E_{f_1,f_2}[-U' + rU' + 100v_1U'']}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''} \\
&= \frac{E_{f_1,f_2}[-(1 - r)U' + 100v_1U'']}{-E_{f_1,f_2}(v_1 + c - v_2)^2 U''}
\end{aligned}$$

The third step utilizes $r < 1$. In the last expression, the numerator is negative as $r < 1, v_1 \geq 0, U'' < 0$ and the denominator is positive, which concludes the proof. $\square$

ONLINE APPENDIX

A Additional tables

Table A.1: The influence of changes in interpersonal uncertainty on ingroup favoritism

	Dependent variable: Allocation to ingroup					
	Both groups pooled		Political views		Hobbies/interests	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant (Ingroup pref.)	0.666^{***}	0.741^{***}	0.964^{***}	0.898^{***}	0.367^{***}	0.584^{***}
	(0.104)	(0.132)	(0.140)	(0.186)	(0.101)	(0.130)
Interpersonal unc.	0.742^{***}		0.720^{***}		0.764^{***}	
	(0.115)		(0.121)		(0.127)	
Higher unc. ingroup		-0.856^{***}		-0.622^{***}		-1.090^{***}
		(0.144)		(0.178)		(0.178)
Higher unc. outgroup		0.629^{***}		0.819^{***}		0.438^{***}
		(0.135)		(0.187)		(0.154)
Subjects	120	120	120	120	120	120
Observations	720	720	360	360	360	360
R^2	0.095	0.096	0.080	0.081	0.121	0.129

Notes: The table shows OLS estimates. The dependent variable is the amount allocated to the ingroup (out of $10), subtracted by five. "Interpersonal uncertainty" is equal to 1, 0, or -1 when the decision has *High uncertainty on ingroup members*, *Group information without uncertainty*, and *High uncertainty on outgroup members* respectively. "Higher uncertainty ingroup" is equal to 1 when the decision has *High uncertainty on ingroup members* (*High uncertainty on outgroup members*) and zero otherwise. In columns (1) and (2), decisions are pooled across both social groups (political views and hobbies/interests) that define the ingroup and outgroup. In columns (3) and (4), only decisions involving groups based on political views are considered, while in columns (5) and (6) only decisions involving groups based on hobbies/interests are considered. Standard errors (in parentheses) are clustered at the subject level. Significance levels: *p$<$0.1, **p$<$0.05 and ***p$<$0.01.

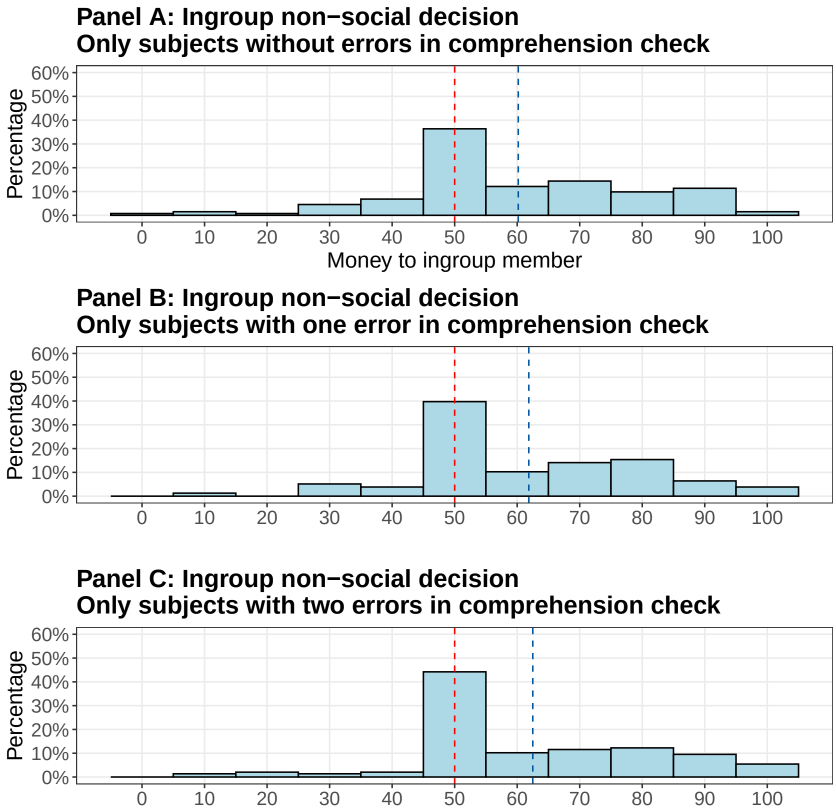
Table A.2: Structural estimation results (Section 3.2.3)

	Main model (1)
γ	0.374***
	(0.016)
b	0.741***
	(0.121)
σ	1.881***
	(0.046)
a_{IU}	0.739***
	(0.152)
a_{ING}	0.336**
	(0.133)
LL	-1718
Akaike's IC	3445
Bayesian IC	3469

Notes: γ*,* b*, and* σ *are the CRRA parameter, measure of ingroup preferences, and the standard deviation of the noise term* ε *respectively.* a_{IU} *and* a_{ING} *quantify the importance of interpersonal uncertainty (IU) and ingroup preferences on the optimal allocation choice when both factors are present. See Section 3.2.3 for details. "LL" denotes the maximized Log-Likelihood, "Akaike's IC" is the Akaike's information criterion and "Bayesian IC" the Bayesian information criterion. Significance levels:* $^{*}p<0.1$*,* $^{**}p<0.05$ *and* $^{***}p<0.01$*.*

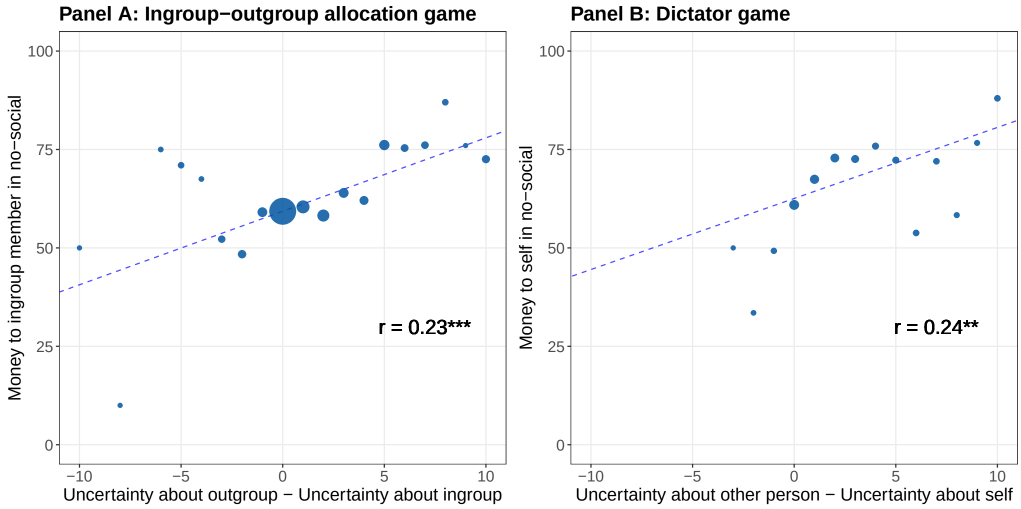
B Additional figures

Figure B.1: Ingroup non-social decisions sorted according to comprehension question errors



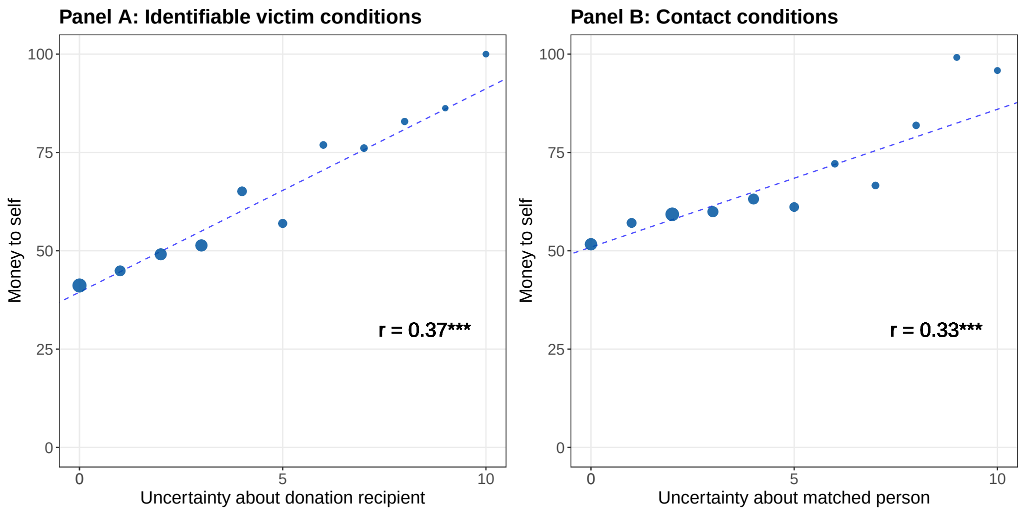
Notes: Histogram of *Ingroup non-social* decisions. The x-axis denotes the amount of money (out of $100) allocated to the ingroup member instead of the outgroup member. The red line denotes the even split, the blue line the average allocation. The decisions have consequences only for the decision-makers, with their payoff depending on the ingroup and outgroup member's valuation of money. In all panels, the binwidth is 10. Decisions are pooled across the three groups (shared hobbies/interests, political views, and religious beliefs).

Figure B.2: Association between reported interpersonal uncertainty and non-social allocation decisions



Notes: Binned scatter plots showing the relationship between reported interpersonal uncertainty and non-social decision-making in ingroup versus outgroup allocation games (Panel A) and the dictator game (Panel B). **Panel A:** The x-axis shows the difference in interpersonal uncertainty between ingroup and outgroup members, and the y-axis shows money allocated to the ingroup member in the *Ingroup non-social* decision. $n = 720$ decisions from 240 subjects, pooled across three social groups and two environments (*Gift card* and *Effort*, see Section 3). **Panel B:** The x-axis shows the difference in interpersonal uncertainty between the dictator game recipient and oneself, and the y-axis shows the amount allocated to self in the *Self non-social* decision. $n = 120$ decisions from 120 subjects. In both panels, we divide the interpersonal uncertainty measure into 20 bins and plot the average allocation in each bin. The linear regression fit is shown with the blue dotted line, and correlation coefficients are displayed. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

Figure B.3: Association between reported interpersonal uncertainty and allocation decisions



Notes: Binned scatter plots showing the relationship between reported interpersonal uncertainty and decision-making for the *Identifiable victim* intervention (Panel A) and the *Contact* intervention (Panel B). **Panel A:** The x-axis shows uncertainty about the recipients of the GiveDirectly donation, and the y-axis shows money allocated to oneself. $n = 304$ decisions from 304 subjects pooled across treatment and control groups. **Panel B:** The x-axis shows uncertainty about the matched person, and the y-axis shows money allocated to oneself. $n = 298$ decisions from 298 subjects pooled across treatment and control groups. In both panels, we divide the interpersonal uncertainty measure into 10 bins and plot the average allocation in each bin. The linear regression fit is shown with the blue dotted line, and correlation coefficients are displayed. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

C Separate results for the Gift card and Effort environment of the Ingroup paradigm

In Section 3.1.2, we pooled our results across the *Gift card* and *Effort* environment. In this Section, we report the results from both environments separately as was preregistered. This allow us to jointly assess whether (i) our results depend on the medium used for allocations and the valuation task (willingness-to-pay for a gift card versus willingness-to-work for bonus payments) and (ii) whether varying the implementation probability influences choices (paying one per session versus paying one out of every ten subjects).

In the following, we report the results from the *Gift card* environment in the text and supply the results from the *Effort* environment in brackets.

***Ingroup social* decisions.** In the *Ingroup social* decisions, subjects allocate on average \$57.48 (\$58.00), \$71.05 (\$70.03), and \$61.61 (\$59.80) to the ingroup member when they share the same interests/hobbies, the same political views, or the same religious beliefs, respectively. In all cases, we can reject the hypothesis of no ingroup-favoritism ($p < 0.01$, one-sample Wilcoxon tests). In 46% (51%) of the decisions, subjects display ingroup-favoritism by allocating strictly more than 50% to the ingroup. Outgroup-favoritism is found in 8% (3%) of decisions, and in the remaining 46% (45%) of cases, subjects allocate 50/50. In total, 73% (73%) of subjects display ingroup-favoritism in at least one decision.

***Ingroup non-social* decisions.** In both environments, a similar pattern emerges in the *Ingroup non-social* decisions. Here, subjects allocate on average \$56.86 (\$60.02), \$65.02 (\$64.20), and \$60.81 (\$60.17) to the ingroup member's WTP when the ingroup member shares the same interests/hobbies, same political views, and same religious beliefs as the subject. As before, we find behavior resembling ingroup-favoritism in all cases ($p < 0.01$, one-sample Wilcoxon tests). In 61% (54%) of the *Ingroup non-social* decisions, subjects display ingroup-favoritism by allocating strictly more than 50% to the ingroup member's WTP. Outgroup-favoritism is found in 11% (9%) of decisions, and in the remaining 28% (37%), subjects allocate 50/50.

Comparing *Ingroup social* and *non-social*. We cannot reject that average ingroup allocations are equal between *Ingroup social* and *non-social* decisions in any case at the 5% level, with p-values of $p = 0.59$ ($p = 0.77$) for hobbies/interests, $p = 0.22$ ($p = 0.06$) for political views, and $p = 0.38$ ($p = 0.68$) for religious beliefs using unpaired Wilcoxon tests.

Within-subject comparison. We also find no substantial difference in the within-subject relationship of *Ingroup social* and *non-social* between the *Gift card* and *Effort* environment. The correlation is highly significant in both cases, with a correlation coefficient of $r = 0.53$ ($r = 0.68$).

Overall, our results are quantitatively very similar across both environments. This suggests that our results are not specific to either the use of gift cards or effort tasks. Moreover, our results are insensitive towards variations in the probability that decisions are implemented with real consequences.

D Rawlsian preferences under interpersonal uncertainty

In Section 2, we showed that utilitarianism generates patterns of prosocial behavior given certain assumptions on interpersonal uncertainty. This raises the question of whether every commonly used welfare criterion delivers similar patterns under the right parameters given our assumptions. Here, we show that Rawlsian preferences – one of the most discussed welfare criterion – are insensitive to interpersonal uncertainty. Under Rawlsian preferences, only the utility of the least well-off recipient matters. In our context, Rawlsian preferences mean the utility individual i receives from allocating x to the ingroup member and $(100-x)$ to the outgroup member is $u_{RAWLS} = \min\{v_1 x, v_2(100-x)\}$. As Theorem 2 shows, a DM will then split the money equally independent of interpersonal uncertainty.

Theorem 2. *Suppose individual i has Rawlsian preferences. Then irrespective of i's risk attitude ($U'' \leq 0$ or $U'' \geq 0$), her optimal allocation is $x^* = 50$, in both the following cases, i) $f_1 = f_2$, and, ii) f_2 is a mean preserving spread of f_1.*

Proof of Theorem 2. As v_1, v_2 are random variables, i's expected utility from allocating x to the outgroup is:

$$EU(x) = E_{v_1 \sim f_1, v_2 \sim f_2} \min\{v_1 x, v_2(100-x)\}$$

For any $x \in [0, 50) \cup (50, 100]$,

$$\min\{v_1(100-x), v_2 x\} + \min\{v_1 x, v_2(100-x)\} \leq v_1(100-x) + v_1 x = 100 v_1$$

with strict inequality whenever $v_1 \neq v_2$.[32]

Similarly, $\min\{v_1(100-x), v_2 x\} + \min\{v_1 x, v_2(100-x)\} \leq 100 v_2$ with strict inequality whenever $v_1 \neq v_2$. Putting these two inequalities together, we get

$$\min\{v_1(100-x), v_2 x\} + \min\{v_1 x, v_2(100-x)\} \quad \leq \quad \min\{100 v_1, 100 v_2\}$$

with strict inequality whenever $v_1 \neq v_2$. Next, using $f_1 = f_2$,

$$\begin{aligned} EU(x) &= E_{v_1 \sim f_1, v_2 \sim f_2} \min\{v_1 x, v_2(100-x)\} \\ &= E_{v_1 \sim f_1, v_2 \sim f_2} \min\{v_1(100-x), v_2 x\} \end{aligned}$$

[32]If $x = 50$, then strict inequality does not hold under $v_1 < v_2$.

Therefore, for any $x \in [0, 50) \cup (50, 100]$,

$$\begin{aligned} EU(x) &= \frac{1}{2} \times E_{v_1 \sim f_1, v_2 \sim f_2} \left(\min\{v_1 x, v_2(100-x) + \min\{v_1(100-x), v_2 x\}\right) \\ &< \frac{1}{2} \times E_{v_1 \sim f_1, v_2 \sim f_2} \min\{100 v_1, 100 v_2\} \\ &= E_{v_1 \sim f_1, v_2 \sim f_2} \min\{50 v_1, 50 v_2\} \end{aligned}$$

The first inequality becomes strict as $v_1 \neq v_2$ with positive probability in the integration. This proves part (i), and a similar proof works for part (ii) after v_2 is replaced with $w_1 + z_1$ like in the proof of Theorem 1.

□

E Within-subject analyses

The results covered in the main text were obtained using a between-subject design, where we only used the first set of decisions each subject faced. In the following, we repeat our analyses using all of the subjects' decisions. In general, our between-subject results replicate well in the within-subject analyses. As in the main text, we pool across the *Gift card* and *Effort* environment.

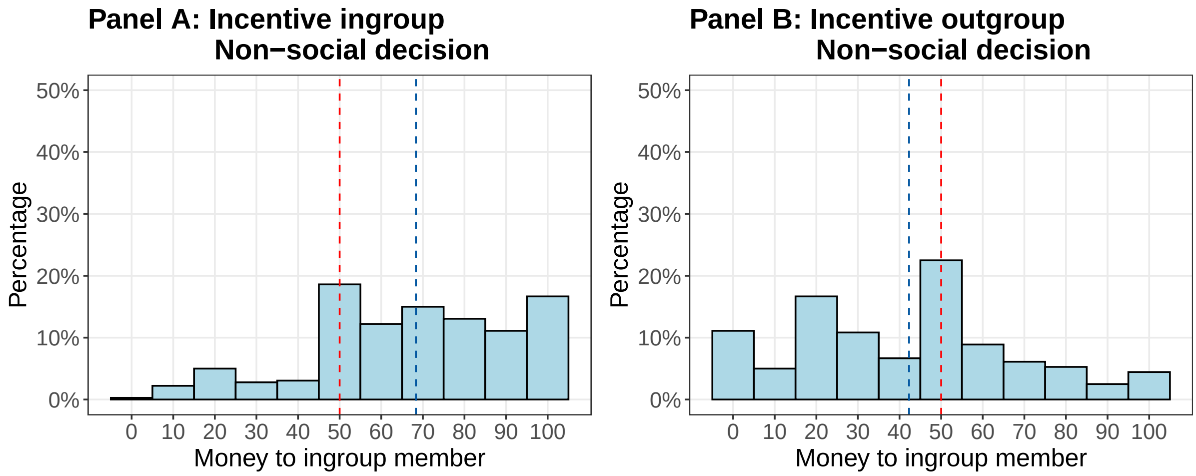
E.1 Ingroup versus outgroup paradigm main results

***Ingroup social* decisions.** In the within-subject case, subjects allocate on average \$58.62 if their ingroup members share the same interests/hobbies, \$68.51 if political views are shared, and \$60.92 if religious beliefs are shared. In all three cases, we can reject the hypothesis of no ingroup favoritism ($p < 0.001$, one-sample Wilcoxon tests). In 52% of the decisions, subjects display ingroup favoritism by allocating strictly more than 50% to the ingroup. Outgroup favoritism is found in 6% of decisions, and in the remaining 42%, subjects allocate 50/50. In total, 75% of subjects display ingroup favoritism in at least one decision.

***Ingroup non-social* decisions.** In the *Ingroup non-social* decisions, subjects allocate on average \$59.08 to their ingroup member's WTP when the ingroup is about shared interests/hobbies, \$63.65 if political views are shared, and \$60.18 if religious beliefs are shared. Again, we can reject the hypothesis of no ingroup favoritism ($p < 0.001$, one-sample Wilcoxon tests) in all three cases. In 54% of the decisions, subjects display ingroup favoritism by allocating strictly more than 50% to the ingroup member's WTP. Outgroup favoritism is found in 11% of decisions, and in the remaining 35%, subjects allocate 50/50.

Comparing *Ingroup social* and *non-social*. Comparing average ingroup allocations between *Ingroup social* and *non-social* within-subject reveals that we cannot reject equality in two of the three cases ($p = 0.77$ for hobbies/interests, $p = 0.003$ for political views, $p = 0.92$ for religious beliefs, paired Wilcoxon tests).

Figure E.1: Ingroup incentive within-subject



Notes: See Figure 3 for details on the variable definitions. Compared to Figure 3, this Figure displays the results using all decisions from subjects in *Incentive Ingroup* and *Incentive Outgroup*. Each panel thus displays $n = 360$ decisions by 120 subjects.

E.2 Ingroup versus outgroup setting incentive robustness

Table E.1 displays the treatment of *Outgroup incentive* relative to *Ingroup incentive* effects separately for the within-subject and between-subject effects pooled across the three groups. As displayed, the effect is similar in both the within- and between-subject comparisons. Regarding the within-subject effects in the social groups individually, when the ingroup is incentivized, average allocations to the ingroup WTP increase from \$58.47 to \$67.22 for hobbies/interests ($p < 0.001$, unpaired Wilcoxon tests), from \$64.00 to \$72.57 for political views ($p < 0.001$), and from \$58.97 to \$65.22 for religious beliefs ($p = 0.01$) compared to *Ingroup non-social*. Conversely, in *outgroup incentive*, allocations to the ingroup WTP decrease to \$37.76 for hobbies/interests, to \$46.89 for politics, and to \$42.21 for religious beliefs (all three $p < 0.001$). As in the between-subject comparison, we again see outgroup favoritism in the *Outgroup incentive* decisions. The pooled average is \$42.29, which is significantly different from the even split ($p < 0.001$, one-sample Wilcoxon test). See Figure E.1 for the distributions, which once again show that the shift in average giving is driven by shifts in the distributions.

E.3 Self versus others setting main results

***Self social* decision.** In the within-subject case of the *Self social* decision, subjects allocate on average \$68.05 to themselves, thus displaying significant self-regarding behavior relative to the equal split ($p < 0.001$, one-sample Wilcoxon test). In total, 62% of subjects allocate more money to themselves, 9% allocate more to the other person, and 29% implement the 50/50 split.

Table E.1: Treatment effect of the incentive treatment in the ingroup versus outgroup setting

	Dependent variable:	
	Allocation to ingroup member	
	Within-subject (1)	Between-subject (2)
Outgroup incentive	-26.047^{***}	-25.817^{***}
	(3.163)	(3.547)
Constant (*Ingroup incentive*)	68.333^{***}	67.978^{***}
	(1.937)	(2.083)
Subjects	120	120
Observations	720	360
R^2	0.211	0.231

Notes: The table shows OLS estimates. The dependent variable in columns (1) and (2) is the amount subjects allocate to themselves (out of $100) in the *Ingroup incentive* and *Ingroup incentive* treatments. “Outgroup incentive” is a dummy equal to one if the incentive for the decision gave three times the weight on the outgroup member’s WTP, and equal to zero if the incentive gave three times the weight on the ingroup member’s WTP. In column (1), all decisions are used, in (2) only the first decisions. Standard errors (in parentheses) are clustered at the subject level. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

***Self non-social* decision.** In the *Self non-social* decision, subjects allocate on average $67.23 to their own WTP, again displaying significant self-regarding behavior ($p < 0.001$, one-sample Wilcoxon tests). In total, 66% of subjects allocate more money to their own WTP, 13% allocate more to the other person’s WTP, and 21% implement the 50/50 split.

Comparing *Self social* and *non-social*. In the between-subject comparison, we also cannot reject equality of average allocations between *Self social* and *non-social* ($p = 0.69$, paired Wilcoxon tests). Similarly, we cannot reject that the pooled distributions are equal ($p = 0.95$, Kolmogorov-Smirnov test).

F Analyzing order effects

A potential concern for the validity of the within-subject results for the *social* and *non-social* decisions (Appendix E) is contagion across conditions. As subjects facing the first set of decisions were not aware that a second set would follow, this naturally cannot influence our between-subject analysis presented in the main paper that only uses the first set of decisions. However, subjects may adjust their choice in the subsequent *non-social* decisions to mimic the *social* decisions, potentially biasing the individual-level analyses. Such adjustment could lead to artificially high similarity between the two decisions, and thus artificially high correlations. Because we randomized the order of decisions, we can directly assess this concern by testing for order effects. Overall, we find no evidence that the order influences subjects' behavior, as we show in the following in detail.

F.1 Ingroup versus outgroup paradigm

For the *Ingroup non-social* decisions, the pooled average allocations to the ingroup are \$61.50 when elicited before, and \$60.45 when elicited after the *social* decisions. For hobbies/interests, *Ingroup non-social* the averages are \$58.61 and \$59.54 ($p = 0.94$, unpaired Wilcoxon test), for political views \$64.60 and \$62.73 ($p = 0.49$), and for religious beliefs \$61.29 and \$59.10 ($p = 0.25$). Thus, the averages are invariant to the order. We also cannot reject the null hypothesis that distributions are invariant to the order ($p = 0.78, p = 0.60, p = 0.39$, Kolmogorov-Smirnov tests). Moving to the *Ingroup social* decisions, average allocations are \$63.05 when *Ingroup social* is elicited first, and \$62.31 when elicited after the *non-social* decisions. Again, decisions generally do not differ substantially. For hobbies/interests *Ingroup social* the averages are \$57.75 and \$59.50 ($p = 0.04$, unpaired Wilcoxon test), for political views \$70.66 and \$66.32 ($p = 0.26$), and for religious beliefs \$60.73 and \$61.12 ($p = 0.21$).

F.2 Self versus other paradigm

In the case of the *Self non-social* decisions, subjects allocate \$64.12 to themselves when the decision is before the *Self social* decision, and \$70.25 when the decision comes afterward, an insignificant difference ($p = 0.12$, unpaired Wilcoxon test). In the case of the *Self social* decisions, subjects allocate \$69.05 to themselves when the decision is before the *Self non-social* decision, and \$67.02 when the decision comes afterward, again an insignificant difference ($p = 0.61$). In addition, we can reject the null that the distributions are invariant to the order at the 5% level for *Self non-social* ($p = 0.08$ Kolmogorov-Smirnov tests) and at any conventional level for *Self social* ($p = 0.54$ Kolmogorov-Smirnov tests).

F.3 Giving versus taking paradigm

In the case of the *Taking non-social* decisions, subjects allocate \$56.00 to themselves when the decision is before the *Taking social* decision, and \$54.05 when the decision comes afterward, an insignificant difference ($p = 0.49$, unpaired Wilcoxon test). In the case of the *Taking social* decisions, subjects allocate \$39.52 to themselves when the decision is before the *Taking non-social* decision, and \$41.51 when the decision comes afterward, again an insignificant difference ($p = 0.52$). In addition, we cannot reject the null that the distributions are invariant to the order both for *Taking non-social* ($p = 0.45$ Kolmogorov-Smirnov tests) and *Taking social* ($p = 0.73$ Kolmogorov-Smirnov tests).

G Details on the robustness treatments of Section 3.1.3

This section provides more details on the results of the robustness treatments that are described in Section 3.1.3.

We start with the between-subject comparison of *Outgroup incentive* versus *Ingroup incentive* by focusing on behavior in the first-assigned choices. We find that subjects respond to changes in the induced utilitarian incentives: Compared to an average giving of \$60.89 to the ingroup members' WTPs across our social groups in *Ingroup non-social* (Section 3.1.2), subjects give on average \$42.16 to the ingroup members' WTPs in *Outgroup incentive* and \$67.98 in *Ingroup incentive*, a significant difference (both comparisons $p < 0.001$, unpaired Wilcoxon test). In particular, behavior switches to outgroup favoritism in *Outgroup incentive*, as the average is significantly smaller than the even split ($p < 0.001$, one-sample Wilcoxon test). Regarding the distributions, 54% of subjects in *Outgroup incentive* and 10% in *Ingroup incentive* display outgroup favoritism, while 33% and 72% display ingroup favoritism, respectively. This significant shift in the distributions ($p < 0.01$, Kolmogorov-Smirnov test) shows that the changes in the averages are not driven by a minority of subjects, but a substantial fraction.

A within-subject analysis reveals that subjects change their behavior in 81% of decisions following the incentive change between the two treatments. Accordingly, in 19% of decisions are subjects unresponsive to changes in the incentives, indicating inattention or confusion. Comparing the behavior in these situations to the main experiment's *Ingroup social*, both average ingroup giving (\$60.84 compared to \$63.38) and the fraction of choices displaying ingroup favoritism (34% of decisions compared to 46%) is lower, while 50/50 splits are more frequent (60% to 46%).

Next, we turn to the *Ingroup non-social minimum* treatment. As predicted, we find that favoritism in either direction is eliminated under the Rawlsian incentive. On average, subjects allocate \$51.31 to the WTP of the individual that shares their interests/hobbies, \$52.85 to the individual that shares their political views and \$49.77 to the individual that shares their religious beliefs. Hence, the treatment did not only significantly reduce ingroup favoritism relative to *Ingroup non-social* (in all three cases $p < 0.01$, unpaired Wilcoxon tests), but eliminated it altogether, as we can no longer reject that average ingroup giving is different from the 50/50 split ($p = 0.31$, $p = 0.13$ and $p = 0.95$ respectively, one-sample Wilcoxon tests). Moreover, the percentage of decisions that implement exactly a 50/50 split increases from 32% in *ingroup non-social* to 58% in *Ingroup non-social minimum*.

These results provide evidence against limited attention or confusion driving our results of the previous section. Changing a single number in the Utilitarian incentive formula reverses the direction of favoritism from ingroup to outgroup favoritism, while moving to Rawlsian incentives eliminates any favoritism. Moreover, in those cases where choices indicate inattention or confusion about the incentives, ingroup favoritism is, if anything, less prevalent relative to the main experiment.

Table G.1: Categorizing open-ended responses that interpret WTW-differences

Category	**Example response**	**Count (% subjects)**
Differences in economic need and motivation	"I think for some people having a higher need for money will motivate them to do more tasks. If they are in more need of the money they may be willing to do it more than someone who does not really need the money."	72 (60%)
Miscellaneous or unclear	"Some may want to work versus some may not want to work."	20 (17%)
Differences in value of time and opportunity cost	"They might value their time different."	16 (13%)
Differences in personality traits	"A lot of people are lazy. Most are okay with doing as minimal amount of work as possible. Other people simply like to ride other people's coattails."	9 (7%)
Differences in disutility from tasks	"People who enjoy this type of work would be more willing to do more work"	4 (3%)

H Interpretation of willingness-to-work differences

Elicitation. We included an open-ended question to capture how subjects interpret differences in our willingness-to-work (WTW) measure, which denotes the number of tasks for which subjects are willing to work to receive a $10 bonus. Specifically, at the end of *Ingroup social* and *Ingroup non-social* experiments in the *Effort environment* – after subjects had completed the WTW elicitation and allocation decisions – they answered the following question:

> Why might some Prolific participants be willing to complete more tasks than other participants in order to receive the $10 bonus payment? Please explain in one to three sentences.

Results. We used the large language model GPT-4o to categorize the open-ended responses from 121 respondents. Results are displayed in Table G.1. The modal subject attributes WTW-differences to economic need and motivation. A small number of respondents mention the closely related channel of differences in the value of time[33], and very few men-

[33]Differences in the opportunity cost of time often reflect underlying differences in economic means or wealth, as individuals with greater financial resources would place a higher value on their time/ leisure and thus be less willing to expend effort.

tion personality traits or the disutility of completing the slider task. Lastly, a fraction of the respondents give ambiguous or hard-to-categorize answers. For robustness, we asked a research assistant to also code the open-ended responses to determine what percentage of respondents mention economic need and motivation as a driver of WTW-differences. This was carried out independently and without knowledge of the GPT-4o classification. They classified 69 (57%) of all respondents as attributing WTW-differences to differences in economic need and motivation.

I Ingroup information decision

In this section, we discuss the design and results of an additional decision that was added to the end of the *Ingroup uncertainty* treatment described in Section 3.2.1.

Design. After making the seven decisions that exogenously informed subjects which group had ingroup affiliation and/or which group had a higher WTW-variance, subjects made an eighth and final allocation decision. In this decision, they were not given any information that distinguished one group from the other ex-ante. They were only informed that the two groups were different: (i) both recipients of one group (which group was unspecified) shared their hobbies/interests (or shared their political views), while both recipients of the other group did not, and (ii) both recipients of one group (which group was unspecified) had a WTW of 12, while the two recipients of the other group had a WTW of 4 and 22. However, ex-ante, they did not know which group had the ingroup members, or which group had a lower variance in WTW, or if ingroup or outgroup members had lower/ higher WTW variance.

Subjects then could choose to learn one of the two dimensions along which the groups differed. That is, they could either learn which group contained only ingroup members and which contained only outgroup members, or, they could learn which group had WTW variation and which had not. We balanced the presentation of both pieces of information by randomizing the order in which the information was introduced and displayed for the choice. Which information subjects choose reveals which factors are of primary importance in their decision process.

Results. We find that 81% of subjects choose to learn about the WTW information, and only 19% choose to learn the ingroup-outgroup information. Accordingly, in our setting, subjects prefer receiving information about interpersonal uncertainty over information about social group membership, indicating its relevance for decisions.

Information choice type analysis. We further use subjects' preference towards receiving information about the recipients' WTWs or group affiliations to validate our type categorization that we develop in Section 3.2.3. We compare the fraction of subjects choosing the

WTW information instead of the group affiliation information across our four behavioral types. In total, 92% of subjects who respond to uncertainty but display no group preference choose the WTW information. This fraction decreases to 78% for those who respond to uncertainty and reveal a group preference, and decreases further to 50% for those who reveal a group preference but do not respond to uncertainty. Hence, our categorization predicts subjects' information choices in the expected direction.

J Self versus others setting incentive treatment

Design. As in the ingroup case, we vary the incentive subjects face when making the *Self non-social* decisions. In *Self incentive*, the weight on the DM's own WTP is three times as high as the other individuals WTP. The DM's payoff thus becomes:

$$\Pi\left(x_{self}, x_{other}\right) = 3 \cdot x_{self} \cdot WTP_{self}/100 + x_{other} \cdot WTP_{other}/100$$

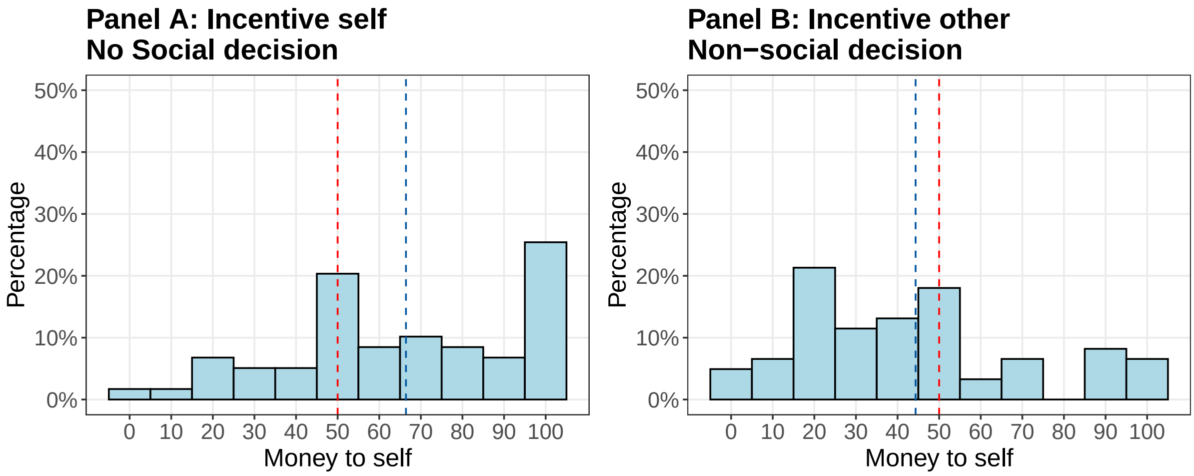
In *Other incentive* we increase the weight put on the other individual's WTP to be three times as high as the DM's WTP:

$$\Pi\left(x_{self}, x_{other}\right) = x_{self} \cdot WTP_{self}/100 + 3 \cdot x_{other} \cdot WTP_{other}/100$$

Results. Inducing these incentives changes people's behavior in the *non-social* decision. See Table J.1 for the within-subject and between-subject treatment effect. In both cases, the change in incentives leads to a significant change in the amount participants allocate to their own WTP. In the within-subject comparison, participants allocate \$19.50 less to their own WTP when incentives are higher for the other participant; in the between-subject comparison, they allocate \$22.08 less. Figure J.1 displays the distributions in the between-subject case. The fraction of subjects allocating more than 50% of the endowment to their own WTP increases from 33% in *Other incentive* to 63% in *Self incentive,* while the fraction of subjects allocating more money to the other participant decreases from 50% to 22%. The distributions are significantly different from each other ($p < 0.001$, Kolmogorov-Smirnov test.

Using the within-subject comparison shows that 83% of subjects change their allocation behavior between *Self incentive* and *Other incentive*. Among those 17% subjects that are unresponsive to the incentive change, 33% allocate more to themselves, a substantially lower fraction than the 58% in the main *Self non-social* case. In total, 38% choose the equal split, and 29% allocate more to the other participant. Taking the behavior of these unresponsive subjects as indicative of inattention or confusion, it appears that such factors are associated with subjects allocating less to themselves. This result thus provides suggestive evidence that our replication of significantly more self-giving in *Self social* using the *Self non-social* decisions is not driven by inattentive or confused subjects.

Figure J.1: Self versus other incentive



Notes: Histogram of *Self incentive* (Panel A) and *Other incentive* (Panel B) decisions. The x-axis denotes the amount of gift card money (out of $100) that subjects allocate to themselves instead of another individual. Subjects incentive is to maximize the weighted sum of their own and another individuals WTP. In Panel A, subjects own WTP receives three times the weight, in Panel B, the other individual's WTP receives three times the weight. The red dotted line denotes the even split benchmark, the blue dotted line the average allocation. For both panels, the binwidth is 10. Only the first decision is used for each subject. Panel A displays $n = 59$ decisions by 59 subjects, Panel B displays $n = 61$ decisions by 61 subjects.

Table J.1: Treatment effect of the incentive treatment in the self versus other setting

	Dependent variable:	
	Allocation to self	
	Within-subject (1)	Between-subject (2)
Other incentive	-19.500^{***}	-22.079^{***}
	(3.468)	(5.045)
Constant (*Self incentive*)	65.625^{***}	66.424^{***}
	(2.563)	(3.592)
Subjects	120	120
Observations	240	120
R^2	0.108	0.140

Notes: The table shows OLS estimates. The dependent variable is the amount subjects allocate to themselves (out of $100) in the *Other incentive* and *Self incentive* treatments. "Other incentive" is an indicator equal to one if the incentive for the decision gave three times the weight on the other person's WTP, and zero if the incentive gave three times the weight on the subject's own WTP. In column (1), all decisions are used, in (2) only the first decisions. Standard errors (in parentheses) are clustered at the subject level. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

K Giving versus taking paradigm

The previous literature primarily finds that redistribution behavior is merit-based: people redistribute less from initial endowments if these endowments are earned compared to generated by chance (Cappelen, Falch, and Tungodden, 2020). In particular, in the context of dictator games, several studies (discussed in the introduction) show that if the initial endowment was earned instead of being windfall, then dictators increase their allocation towards the individual earning the endowment. This behavior is typically attributed to fairness preferences (e.g., Tungodden and Cappelen, 2019), fairness-based social norms (Krupka and Weber, 2013), or the role of property rights (Oxoby and Spraggon, 2008).

Interpersonal uncertainty offers a belief-based alternative explanation: if DMs believe that on average, the recipient's disutility from losing earned money exceeds their utility from gaining money (i.e., a gain-loss asymmetry), then DMs would perceive *mean-shifted* interpersonal uncertainty (see Definition 1) for recipients who have earned the endowment compared to recipients who have not. Thus, as suggested by result (iii) of Theorem 1, a simple utilitarian motive under uncertainty would also lead to the same asymmetry between giving and taking environments. Our next treatments test this channel of mean-shifted interpersonal uncertainty.

K.1 Experimental design

Following the typical setup of the literature, we alter our dictator game from a giving environment to a taking environment. DMs face a *Taking social* decision and a *Taking non-social* decision. In both *Taking* decisions, DMs are matched to a previous participant who has earned \$100 for participating in a previous study, scheduled to be paid in 6 weeks from the study day. In total, 123 subjects participated in this experiment.

***Taking social* decision.** In the *social* variant, the DM decides whether to take some or all of the money that the other participant has earned for themselves, adapting the design of Oxoby and Spraggon (2008). The chosen allocation is then implemented with consequences for the DM and the other participant.

***Taking non-social* decision.** In the *non-social* variant, we replicate the setup described in Section 3 with one key difference: because the other participant already earned the \$100 that was up for splitting, the DM's utilitarian incentives were calculated using the other participant's willingness-to-accept (WTA) for gift card money, instead of their WTP. Thus the DM's payment depended on their own WTP and the matched participant's WTA. Specifically, the incentive for the DM is as follows:

$$\Pi\left(x_{self}, x_{other}\right) = x_{self} \cdot WTP_{self}/100 + x_{other} \cdot WTA_{other}/100$$

with x_{self} and x_{other} denoting the money DMs allocate to themselves and the other individual respectively, WTP_{self} is their own WTP and WTA_{other} is the other individual's WTA for the gift card money.

After the DMs participated in the MPL that elicits their WTP, we explained to them the following details about matched participants: First, the matched participants earned the $100 gift card through their participation. Then, we asked them whether they would be willing to give away the gift card in exchange for an immediately payable monetary amount. We ask this question for different amounts of the immediately payable money, using a MPL, to elicit their WTA. The DMs are already familiar with the MPL-elicitation method at this point. We emphasize to DMs that the only difference between their's and the matched participant's elicitation is, instead of having the option to receive the gift card, the matched participants already 'owned' the gift card and had the opportunity to sell it.

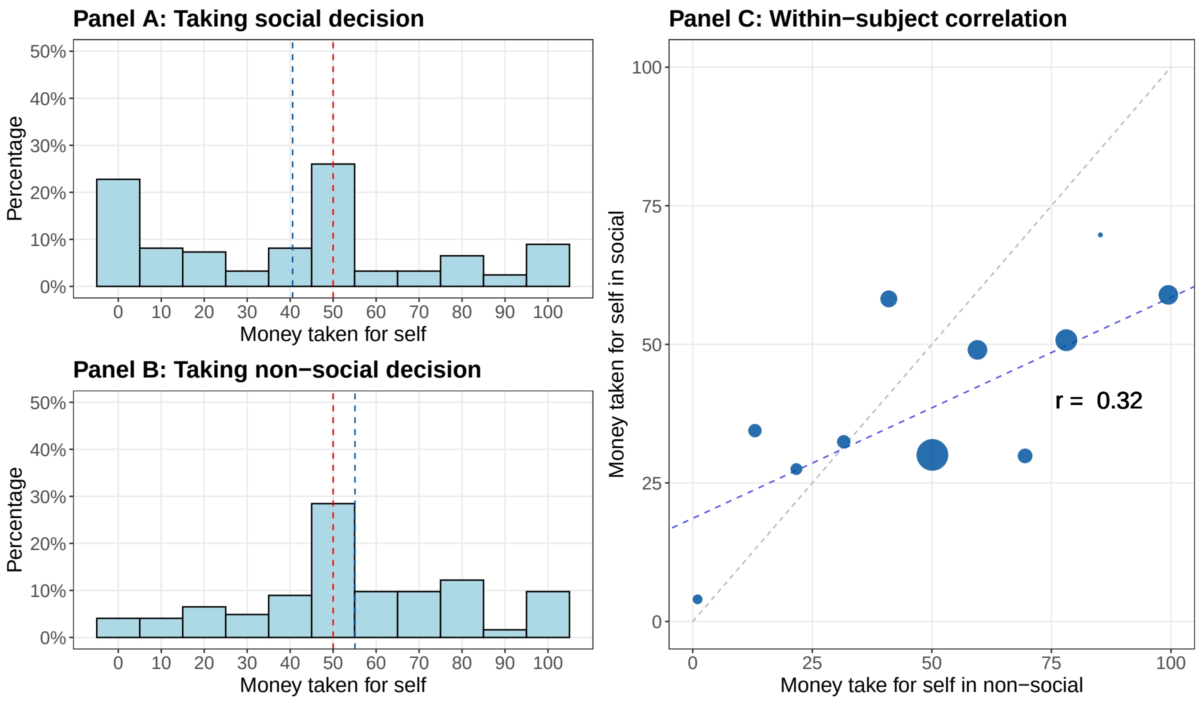
Multiple studies have found that WTA is, on average, higher than WTP (e.g., Camerer, 1995), and hence WTA>WTP is a well-established empirical pattern. Our central hypothesis is that, if DMs also anticipate the WTA-WTP gap as mean-shifted interpersonal uncertainty, then utilitarianism provides a novel foundation for differences in giving and taking (Theorem 1). In particular, under WTA>WTP, we predict that the amount allocated to the matched participant should increase (compared to the giving paradigm) not only in *Taking social*, but also in *Taking non-social*. Further, because *Taking non-social* does not feature any scope for fairness attitudes, we can separate our channel from a fairness channel.

K.2 Results

Comparing *Taking* to the *Giving* setting. We find that subjects allocate significantly less money to themselves in the *Taking* compared to the *Giving* setting of Section 4. Comparing *Self social* with *Taking social* decisions, we see a significant decrease of $27.48 in the amount subjects allocate to themselves, using the within-subject data controlling for the order (see Appendix Table K.1 column (1) for details). We thus replicate the common finding of aversion to taking from earned endowments in the literature with our *social* decisions. When comparing *Self non-social* with *Taking non-social* decisions, we also find a significant decrease of $12.08 in the amount subjects allocate to themselves (Appendix Table K.1 column (2)). Therefore, incentivizing DMs with the other individuals' WTA instead of their WTP induces DMs to allocate less to themselves. Yet, the decrease in allocation to the self from the *Giving* to the *Taking* setting is smaller in the *non-social* case compared to the *social* case.

Comparing *Taking social* and *Taking non-social*. Figure K.1 displays the comparison. Panel A shows the distribution of choices in the *Taking social* decision, where 26% of subjects allocate more money to themselves, 51% allocate more to the other person, with the remaining 23% allocating the even split. In the *Taking non-social* decision, displayed in Panel

Figure K.1: Giving versus taking results



Notes: **Panel A and B:** Histogram of the *Taking social* (Panel A) and *Taking non-social* (Panel B) decision. The binwidth is 10. The x-axis denotes the amount of gift card money (out of \$100) that subjects allocate to themselves instead of another individual. The red line denotes the even split, the blue line the average allocation. In Panel A, subjects' decision has consequences for themselves and the other individual. In Panel B, the decision has consequences only for the subjects, with their payoff depending on their WTP and the other individual's WTA for the gift card. **Panel C:** Binned scatter plot of the *Taking social* and *Taking non-social* decision. We divide *Taking non-social* choices into 10 bins and plot the average giving to self in *Taking social* in each bin. The blue line displays the linear fit of regressing *Taking social* on the *Taking non-social* decision. The correlation coefficient is $r = 0.32$. Displayed are $n = 123$ decision-pairs by 123 subjects.

B, 46% allocate more money to themselves, 29% allocate more to the other person, and 24% split evenly. We see a significant within-subject correlation of $r = 0.32$ ($p < 0.001$) between *Taking social* and *Taking non-social* (Panel C of Figure K.1). Thus, taking behavior in the *social* decision correlates with the *non-social* decision that does not feature taking (not even in how the instructions were framed). Contrary to the other settings, these decisions differ in average allocations ($p < 0.001$, unpaired Wilcoxon test), and distributions ($p < 0.01$, Kolmogorov-Smirnov test). These results suggest that *social* decisions are also driven by motives that are absent in the *non-social* decisions.

A potential motive comes from the observation that 22% of subjects choose to take \$0 for themselves in *Taking social*, while only 3% do so in *Taking non-social*. In contrast, in *Self social* and *Self non-social*, not a single subject chooses to give everything to the other individual. This pattern suggests that fairness preferences are also at work, e.g., some subjects have a strong libertarian fairness view (Alma, Cappelen, and Tungodden, 2020) or adhere to a deontological motive that it is not permissible to take money from someone (Bénabou, Falk, and Henkel, 2024). Interestingly, those subjects refusing to take any money completely

Table K.1: Dictator game allocations under giving setting versus taking setting

	Dependent variable:	
	Allocation to self	
	Social decision	*Non-social decision*
	(1)	(2)
Constant (Giving setting)	68.050***	66.193***
	(2.719)	(2.634)
Taking setting	−27.481***	−12.077***
	(3.529)	(3.114)
Order: *social* decision first	−0.001	2.046
	(3.545)	(3.114)
Observations	243	243
R^2	0.201	0.061

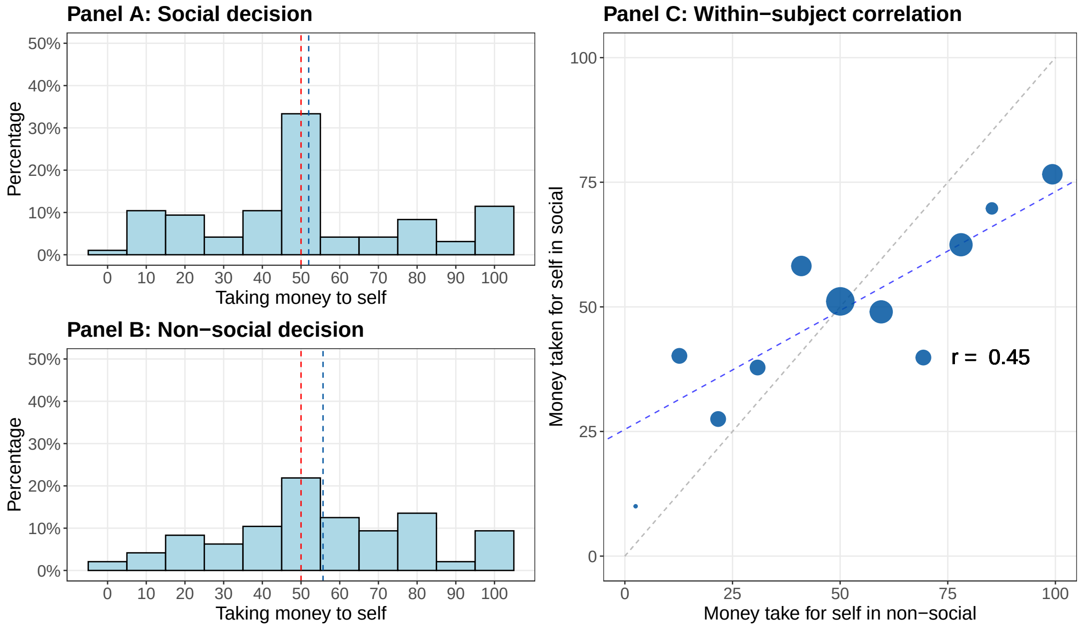
Notes: The table shows OLS estimates. The dependent variable is the amount subjects allocate to themselves (out of $100) in the *Self social* (1) and *Self non-social* decision (2). "Taking setting" is an indicator equal to one if the decision means taking earned money away. "Order: *social* decision first" is an indicator equal to one if subjects faced the *social* before the *non-social* decision. Robust standard errors in parentheses. Significance levels: *p<0.1, **p<0.05 and ***p<0.01.

explain the gap between *Taking social* and *Taking non-social*. If we focus only on subjects who take more than $0 for themselves in *Taking social*, we can no longer reject the equality of average giving between *Taking social* and *Taking non-social* ($p = 0.22$, unpaired Wilcoxon test)[34] and distributions become more similar ($p = 0.09$, Kolmogorov-Smirnov test), see Appendix Figure K.2. Similarly, the within-subject correlation increases to $r = 0.45$.

Beliefs about WTA-WTP differences. A necessary condition our hypotheses and results is that subjects indeed perceive a positive difference in the utility impact of taking earned money and giving windfall money, i.e., in other's WTA and WTP. To validate this assumption, we asked subjects whether they generally think that a person's WTA for the gift card is higher, lower or equal to the WTP. In total, 46% of subjects believe WTA to be higher than WTP, 29% believe WTP to be higher, and 24% believe both to be equal. Thus, subjects believe WTA>WTP on average. For robustness, in the *IU belief measurement* treatment (for details, see the next section), we directly elicited subject's beliefs about both the average WTP and WTA valuations of others. On average, subjects reported a $8.00 higher WTA than WTP ($91.43 compared to $83.43), a significant difference ($p < 0.001$, paired Wilcoxon-test). In total, for 64% of subjects their WTA estimate is higher than the WTP estimate, for 20% the reverse holds, and for 16% both are equal.

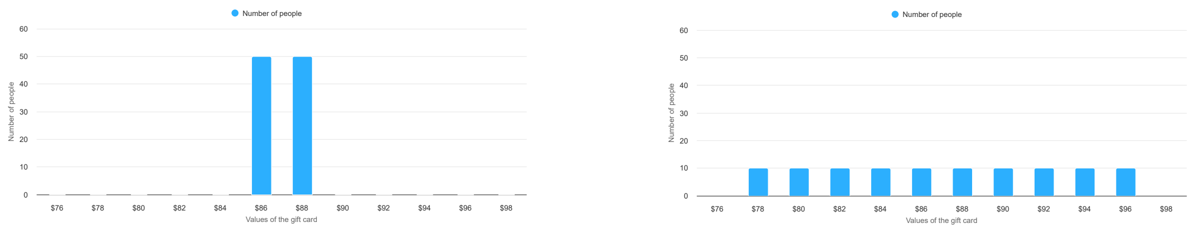
[34]Note that this effect is not mechanical because we remove both *social* and *non-social* decisions due to the within-subject structure of our data.

Figure K.2: Giving versus taking results excluding non-takers



Notes: See Figure K.1 for details on the variables. Compared to Figure K.1, we excluded subjects that take nothing from the other individual in *Taking social*. Thus, displayed are $n = 96$ decisions by 96 subjects.

Figure L.1: Distributions shown to subjects



(a) Distribution with low variance

(b) Distribution with high variance

L Validation of the uncertainty measure

Our self-reported interpersonal uncertainty measure is intended to proxy whether subjects perceive higher interpersonal uncertainty of one group over another, as defined in Definition 1 in Section 2. However, it could be the case that subjects instead only report their perception about mean differences or concepts unrelated to uncertainty. In this section, we validate that our measure is indeed sensitive to those changes in interpersonal uncertainty captured by our definition. To do so, we provide subjects with two objective WTP distributions, one being a mean-preserving spread of the other, and investigate the impact on the answers subjects give to our measure.

Design. At the end of the *Ingroup belief measurement* and *Self/other belief measurement* treatments, we showed subjects two figures. Each figure displayed a frequency distribution of the WTP values of 100 fictitious individuals. In one, 50 individuals had a WTP of \$86 and 50 a WTP of \$88 (low variance distribution). In the other were 10 individuals for each of the 10 values between \$78 and \$96 (high variance distribution). See Figure L.1 for the figures shown to subjects. We also provided these values to subjects in text format below the figures. The high variance distribution is a mean-preserving spread of the low variance one, having the same mean but a lower variance. For each figure, subjects were asked the following about the group displayed in the figure: "Suppose we randomly pick one of the 100 people from this group. How certain are you about how much the randomly chosen person would value the Amazon gift card money?" Subjects could respond on an 11-point Likert scale from *Very uncertain* to *Very certain*, and we re-code the variable so that higher values indicate higher uncertainty. The text and measurement thus closely mirror our self-reported interpersonal uncertainty measure.

Results. We find that subjects report different uncertainty across the two distributions. On average, they report an uncertainty of 3.46 Likert-scale points for the low variance distribution, and an uncertainty of 5.80 points for the high variance distribution, a significant difference ($p < 0.001$, paired Wilcoxon-test). On the individual level, 74% of subjects report a higher uncertainty for the high variance distribution compared to the low variance distribution, 13% report no difference, and the remaining 13% report more uncertainty for the low variance distribution. Thus, subjects are sensitive to changes in WTP distributions in the expected direction.

M Interactions of subjective and objective uncertainty

Researchers have used objective uncertainty to experimentally study excuse-driven selfish behavior. In this section, we extend our simple model to incorporate objective risk–applied to the choice environment studied in Exley (2016)–to demonstrate how interpersonal uncertainty alone can generate behavior that mimics excuse-driven selfishness.

Exley (2016) finds that individuals are more tolerant of risk in their own payoffs than in the payoffs to others, leading to a lower willingness to give when recipients face uncertainty. She interprets this asymmetric risk response pattern as evidence that people use uncertainty as an "excuse" to behave selfishly, motivated by self-image concerns. In this section, we propose a distinct and novel mechanism: even in the absence of excuse-driven motives, such asymmetries can emerge naturally from DMs being averse to the interpersonal uncertainty about others.

M.1 An alternative mechanism for asymmetric responses to objective uncertainty

Similar to Section 2, we assume that the DM values her own payoff as $u_{\text{self}}(x) = v_S x$ where $\$x$ is the amount allocated to herself and v_O is her per-dollar utility. She believes the other's valuation is $u_{\text{other}}(x) = v_O x$, where v_O is unknown and distributed as $\mathcal{F}$ on $[v_{O,l}, v_{O,h}]$. The DM aggregates joint outcomes using a utilitarian aggregator, $U(z)$. For simplicity, we assume that there is no uncertainty about v_O and $E_{\mathcal{F}} v_O = v_S$. Our results would also hold under the more general condition that v_S is a mean preserving spread of v_O.

For any monetary allocation involving risk, the DM evaluates the expected utility over the random component, including the uncertainty about the other's valuation. Valuations of self-lotteries are denoted as $Y_s^j(p)$, and valuations of other lotteries are denoted as $Y_o^j(p)$. The superscript j indicates whether lottery valuations are in self-dollars ($j = s$) or other-dollars to the other person ($j = o$). Self-dollar valuations represent dollars given to participants, while other-dollar valuations represent dollars given to others.

Certainty equivalent definitions. Let $p \in [0, 1]$ denote the probability of a probabilistic outcome (e.g., a lottery). We define four certainty equivalents based on indifference conditions:

Self vs. risky gift to other: $Y_o^s(p)$. Let Y_o^s denote the amount the DM must receive for sure such that she is indifferent to giving the other person \$10 with probability p. Formally,

$$U(v_S \cdot Y_o^s) = p \cdot \mathbb{E}[U(10 \cdot v_O)].$$

Other vs. risky gain to self: $Y_s^o(p)$. Let Y_s^o be the amount the DM is willing to give to the other for sure in place of receiving \$10 with probability p. This satisfies:

$$\mathbb{E}[U(v_O \cdot Y_s^o)] = p \cdot U(v_S \cdot 10),$$

Self vs. risky gain to self: $Y_s^s(p)$. Let Y_s^s be the amount the DM must receive for sure to be indifferent to receiving \$10 with probability p, with no uncertainty. Then:

$$U(v_S \cdot Y_s^s) = p \cdot U(v_S \cdot 10)$$

Other vs. risky gift to other: $Y_o^o(p)$. Let Y_o^o denote the amount the DM is willing to give for sure to the other in place of giving \$10 with probability p. Then:

$$\mathbb{E}[U(v_O \cdot Y_o^o)] = p \cdot \mathbb{E}[U(v_O \cdot 10)]$$

Given the certainty equivalents, we have the following result:

Result 3. *Suppose $E_{\mathcal{F}} v_O = v_S$ and $U(0) = 0$. Then, $U'' > 0$ implies $Y_s^o > Y_o^s$ for $p \in (0, 1)$. If $U(z) = z^r$, with $r \neq 0$ and $\mathcal{F}$ is uniform, then $Y_o^o = Y_s^s$.*

The first part of the result holds more generally, whereas the second part requires assuming the popular CRRA or power functional form.

Proof. The asymmetry between Y_s^o and Y_o^s is most clearly captured through the following chain of inequalities:

$$U(v_S Y_o^s) = p \cdot \mathbb{E}[U(10v_O)] < p \cdot U(10v_S) = \mathbb{E}[U(v_O Y_s^o)] < U(v_S Y_s^o)$$

where the first and last inequalities follow from Jensen's inequality, and the equalities come from the definitions of Y_o^s and Y_s^o respectively. When one further assumes U to be CRRA, then

$$\begin{aligned} U(v_S Y_s^s) &= p \cdot U(v_S \cdot 10) \\ (v_S Y_s^s)^r &= p \cdot (v_S \cdot 10)^r \\ Y_s^s &= 10 \cdot p^{1/r} \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{E}[U(v_O Y_o^o)] &= p \cdot \mathbb{E}[U(v_O \cdot 10)] \\ Y_o^{or} \cdot \mathbb{E}[v_O^r] &= p \cdot 10^r \cdot \mathbb{E}[v_O^r] \\ Y_o^o &= 10 \cdot p^{1/r} \end{aligned}$$

Thus, we have:

$$Y_s^s(p) = Y_o^o(p) = 10 \cdot p^{1/r}$$

□

N Research transparency

All experiments covered in the paper were preregistered at aspredicted.org. The preregistrations include details on the experimental design, the planned sample size, exclusion criteria, hypotheses, and the main analyses. Table N.1 provides an overview over the treatments and links to the respective pre-registrations.

Our experimental implementation followed closely the pre-registration. In particular, we implemented the experimental design and sample size exactly as specified in the pre-registration. Similarly, we employed the exclusion criteria as pre-registered: we specified to exclude any subject who did not complete the experiment. This lead to the exclusion of 22 subjects in the *Ingroup social* and *Ingroup non-social* treatments, 28 in the *Ingroup incentive* and *Outgroup incentive*, 6 in the *Ingroup minimum*, 13 in the *Ingroup effort social* and *Ingroup effort non-social*, 15 in *Ingroup uncertainty*, 22 in *Self social* and *Self non-social*, 23 in *Self incentive* and *Other incentive*, 25 in *Self taking social* and *Self taking non-social*, 4 in *IU belief measurement* and 6 in the *Contact* and *Identifiable victim* interventions. The sample sizes reported in Table N.1 are the final sample sizes used in all analyses of the paper after excluding the previously mentioned numbers of subjects.

Table N.1: Overview over treatments

Label	N	Covered in	Link to preregistration
Ingroup social & *Ingroup non-social Gift card environment*	119	Section 3	https://aspredicted.org/H81_KQ5
Ingroup effort social & *Ingroup effort non-social Effort environment*	121	Section 3	https://aspredicted.org/53G_PNJ
Ingroup incentive & *Outgroup incentive*	120	Section 3.1.3	https://aspredicted.org/H81_KQ5
Ingroup minimum	62	Section 3.1.3	https://aspredicted.org/J7H_W8R
Ingroup uncertainty	120	Section 3.2	https://aspredicted.org/53G_PNJ
Self social & *Self non-social*	120	Section 4	https://aspredicted.org/ZMF_CD9
Self incentive & *Other incentive*	120	App. Section J	https://aspredicted.org/ZMF_CD9
IU belief measurement	120	Sections K.2, 5	https://aspredicted.org/T7X_747
Contact and *Identifiable victim* interventions	602	Section 5	https://aspredicted.org/37tq-7skc.pdf
Self taking social & *Self taking non-social*	123	App. Section K	https://aspredicted.org/RT4_TQB

N.1 Deviations from the pre-registration

The *Gift card* and *Effort* environments of the *Ingroup social* and *Ingroup non-social* treatments were pre-registered and run separately (the *Effort* environment as a robustness check for the *Gift card* environment). In the main text, we pool both environments together. In Appendix Section C, we report the separate results, as pre-registered. In addition, the pre-registrations for the *Ingroup social* and *Ingroup non-social Gift card* treatments as well as the *Self social* and *Self non-social* treatments contain another set of treatments labeled *Group info* and *Self info*. These treatments are not part of this paper and their results are available upon request because the design is superseded by the *Ingroup uncertainty* experiment.[35] The analyses contained in Section 5.1 were pre-registered as exploratory analyses. The structural model of Section 3.2.3 was not pre-registered.

O Experimental instructions

The instructions of all experiments can be found in the following Open Science Framework (OSF) repository:
https://osf.io/tcp3d/?view_only=27899531990048e4b608d64b13528236

[35]The omitted treatments show that providing subjects with information on the WTP of the recipients significantly changes their allocation behavior both in *Ingroup social* and *Self social*. However, in contrast to the *Ingroup uncertainty* experiment, this information manipulation does not directly manipulate interpersonal uncertainty and is potentially confounded by experimenter demand effects.