Journal of Economic Psychology 98 (2023) 102654

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Self-serving bias in redistribution choices: Accounting for beliefs and norms

Dianna R. Amasino [a,*], Davide Domenico Pace [b,a,c], Joël van der Weele [a,c]

[a] CREED, Amsterdam School of Economics, University of Amsterdam, The Netherlands
[b] Ludwig-Maximilians-Universität München, Germany
[c] Tinbergen Institute, The Netherlands

ARTICLE INFO

Dataset link: https://osf.io/qjy6u/files/osfstorage

JEL classification:
C91
D63
D83

Keywords:
Redistribution
Self-serving bias
Fairness
Norms
Online experiments

ABSTRACT

We explore the psychological mechanisms underlying self-serving redistribution decisions in an experimental setting. This self-serving bias in redistribution has been attributed not only to self-interest, but also to constructs such as differing beliefs about the hard work or luck underlying inequality, differing fairness views, and differing perceptions of social norms. In this study, we directly measure each of these potential mechanisms and compare their mediating roles in the relationship between status and redistribution. In our experiment, participants complete real-effort tasks and then are randomly assigned a high or low pay rate per correct answer to exogenously induce (dis)advantaged status. Participants are then paired and those assigned the role of dictator decide how to divide their joint earnings. We find that advantaged dictators keep more for themselves than disadvantaged dictators and report different fairness views and beliefs about task performance, but not different perceptions of social norms. Further, only fairness views play a significant mediating role between status and allocation differences, suggesting this is the primary mechanism underlying self-serving differences in support for redistribution.

1. Introduction

People with higher incomes often support less redistribution than those with lower incomes, a finding that has been consistently shown across surveys, field, and lab experiments (Cohn et al., 2019; Di Tella et al., 2007; Konow, 2000; Koo et al., 2023; Suhay et al., 2021). This gap in support for redistribution could be due purely to self-interest. However, in line with self-image and reputational motivations to appear moral to oneself or others, people often do not go to selfish extremes. Instead, they find excuses or justifications that allow them to support fairness ideals that most benefit themselves. This is especially pernicious in privileged or powerful individuals who are in a position to institutionalize their self-serving bias,[1] which has been linked to polarization, resentment, and social conflict in Western democracies (Babcock et al., 1995; Piketty, 2020; Sandel, 2020; Schwardmann et al., 2022).

While self-serving redistribution decisions are well-documented, their psychological antecedents are less well understood. Theories of fairness and cognitive dissonance have invoked various psychological pathways, including shifts in personal fairness views (Konow, 2000), biased perceptions of social norms (Bicchieri et al., 2023), or motivated beliefs about merit and returns to

* Correspondence to: Center for Experimental Economics and Political Decision-Making (CREED), Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.
E-mail address: d.r.amasino@uva.nl (D.R. Amasino).

[1] Note: our definition of self-serving bias is self-serving judgments of a fair division (Cappelen et al., 2007; Rodriguez-Lara & Moreno-Garrido, 2012). These self-serving biases are different than the common definition of self-serving attribution bias in social psychology, which means to attribute good outcomes to one's ability or effort while attributing bad outcomes to external circumstances such as bad luck (Bradley, 1978; Deffains et al., 2016; Dorin et al., 2021; Miller & Ross, 1975).

https://doi.org/10.1016/j.joep.2023.102654
Received 7 February 2023; Received in revised form 3 June 2023; Accepted 6 July 2023
Available online 13 July 2023

effort (Bénabou & Tirole, 2006; Deffains et al., 2016). Many empirical papers have looked at the role of individual psychological constructs, but there are few comparisons of their relative importance. Moreover, error in the measurement of these constructs has complicated the effort to quantify their explanatory power.

In this paper, we directly measure and investigate the role of these three psychological constructs in redistribution decisions, examining how each construct is affected by status and its potential mediating role in the effect of status on redistribution decisions. First, we look at "personal norms" that characterize what people regard as fair. Personal norms reflect privately held views of fairness that develop out of experience and moral reasoning. They are predictive of pro-social or selfish behavior in economic allocation decisions (Bašić & Verrina, 2021; Messick & Sentis, 1979). Second, we look at "social norms", that is, people's perceptions of what others think is fair. In our setting, social norms are determined by beliefs about which fairness principle(s) most people endorse. The desire to conform with others' views makes social norms predictive of individuals' actions (Krupka & Weber, 2013).

While personal and social norms often align, they are different constructs and can diverge in meaningful ways. For example, most young, married men in Saudi Arabia privately support women working outside the home. Still, a presumed social norm against women's labor force participation undermines support for their wives' job searches (Bursztyn et al., 2020). In the context of climate change, Andre et al. (2021) and Sparkman et al. (2022) find that most Americans are willing to support mitigation efforts, but they underestimate others' support for mitigation, undermining collective action. Findings from experimental data on allocation decisions also suggest that these constructs have separate predictive power for behavior (Bašić & Verrina, 2021). Moreover, the extent to which different constructs predict behavior may depend on the strength of social image concerns and expectations of conformity (Ajzen & Fishbein, 1970; Bašić & Verrina, 2021; Cialdini et al., 1991; Thøgersen, 2008).

Third, we consider beliefs about the determinants of economic success and inequalities. High-income people are more likely to attribute their success to hard work and ability than luck (Cassar & Klein, 2019; Deffains et al., 2016; Di Tella et al., 2007; Dorin et al., 2021; Suhay et al., 2021; Valero, 2021). In contrast, those who are less successful or experience hardship are more likely to point to the role of luck or selfishness in success (Almås et al., 2022; Hochleitner, 2022; Hvidberg et al., 2020). Beliefs about the determinants of success have been shown to influence people's preferences for redistribution, as people are more likely to redress inequalities due to luck rather than differences in effort (e.g. Bortolotti et al., 2017; Cappelen et al., 2013, 2023; Cherry et al., 2002; Durante et al., 2014; Krawczyk, 2010; Lefgren et al., 2016).

In this study, we investigate with an experiment how having a privileged status impacts these three constructs, and we study their mediating role in allocation decisions. We do so in the context of a large online experiment with a sample of 600 participants based on the design of Konow (2000). In the experiment, participants first work on real-effort tasks to produce earnings. We manipulate status by randomly assigning half of the participants a higher pay rate per correct answer in the tasks, such that half have a pay advantage and half have a pay disadvantage. Participants then act as "dictators" deciding how to divide joint task earnings, first between themselves and another participant and then between two others in which they have no stake of their own. In this setting, we replicate the findings of Konow (2000), who showed that participants advantaged by a randomly-assigned higher pay rate keep more of the joint earnings and continue to favor other advantaged workers even when self-interest is removed. This persistence to "impartial" decisions is particularly indicative of self-serving bias, and it is the focus of our investigation.

Our original contribution is in (1) examining how the randomly-assigned (dis)advantage in pay rate (or 'status') impacts personal norms, social norms, and beliefs and (2) quantifying and comparing the mediating roles of each construct in the relationship between status and divisions of joint earnings accounting for measurement error. We find that status differences lead to self-serving shifts in personal norms and beliefs, but we find no statistically significant effect for social norms. Moreover, participants show awareness of the bias induced by status in fairness principles when predicting others' norms. Finally, we show that differences in divisions of joint earnings due to dis(advantaged) status in impartial decisions are primarily mediated by shifts in personal norms, with minimal contributions of social norms and beliefs. This result points to a primary role of shifting personal norms (without significant changes in perceptions about what others find appropriate) in driving self-serving attitudes toward redistribution.

Our findings go beyond existing empirical work that either infers psychological mechanisms from shifts in behavior or focuses on a particular mechanism. Konow (2000) and Rodriguez-Lara and Moreno-Garrido (2012) found that participants who benefit from luck incorporate it into their fairness principle when dividing joint earnings, supporting the idea that personal norms adapt to the context. However, they do not explicitly measure personal norms, social norms, or beliefs – they infer this from allocation choices. Deffains et al. (2016) identify self-serving biases in the selection of redistribution criteria as well as a corresponding shift in attribution whereby more successful dictators are more likely to attribute their success to effort. However, they do not explicitly study the link between these variables. Dorin et al. (2021) use the setup of Deffains et al. (2016) to explore the role of in-group bias and personal norms as mediators of self-serving biases, finding that both act as contributing mechanisms of the bias. Lobeck (2023) and Valero (2021) show that participants distort beliefs about performance independently of monetary incentives to do so. Yet, they do not quantify the mediating role of beliefs in self-serving biases. Ubeda (2014) runs a descriptive study where she classifies the dictators' fairness norms.

2. Theoretical framework

Our introduction cites work showing that socio-economic status affects beliefs about fairness and merit and attitudes towards redistribution. To explain these observations, several papers have invoked concepts like cognitive dissonance (Konow, 2000) or motivated reasoning (Suhay et al., 2021). According to such accounts, the wish to justify the status quo and limit redistribution to the less fortunate leads people to self-servingly manipulate their fairness ideals and attributions of success. In Supplementary section B, we formalize this idea in a model inspired by Cappelen et al. (2007). The model captures a simple division problem –

mirroring the setup of the current experiment and earlier experiments – where a decision maker allocates a sum of money that has been produced by herself and another person. Crucially, one of the two agents randomly receives a relative "advantage" in the production process, whereby her performance is multiplied by a higher pay rate, boosting her production share in the total surplus to be divided.

When dividing the surplus, we assume that decision-makers care both about their own payoff and about the fairness of the allocation. Specifically, we assume they adhere to one of several fairness criteria that have been identified in the literature (Cappelen et al., 2007; Konow, 2000; Rodriguez-Lara & Moreno-Garrido, 2012): egalitarian (equal split), meritocratic (proportional to task performance), and libertarian (proportional to the share of total surplus produced — i.e., including randomly determined pay rate advantage). As fairness is subjective, agents may differ in which fairness criterion they deem most appropriate, or they may put some weight on all criteria. If the chosen allocation differs from their subjective fairness ideal, decision-makers incur a psychological cost in terms of self-image or guilt.

Thus, decision-makers in the model navigate a trade-off between taking more money for themselves and remaining closer to their subjective fairness ideal. This trade-off generates pressure to shift their subjective fairness ideal in a self-serving direction to increase the amount they can allocate to themselves without increasing guilt. As an example, consider an advantaged subject in the role of dictator. Because of her advantage, she will typically outperform the recipient in terms of the total contribution, although not necessarily on the "raw" task performance. This implies that the libertarian fairness criterion will be the most advantageous, as it prescribes taking a high share for herself.

We expand the model to capture the cognitive channels responsible for such self-serving bias. We assume that decision-makers may shift their weights on the different fairness criteria, as a function of their advantaged status. They can do so by changing their personal and social norms as well as the attributions of success. In terms of our example, we assume the advantaged decision maker may convince herself (a) that the libertarian criterion is the most appropriate one (personal norms), (b) that this view is generally shared among other participants so that she would find support for her decisions by others (social norms), and (c) that her relative performance is higher than it actually is, so that she is entitled to a bigger share. In the model, these processes will increase the weight on the libertarian criterion in her fairness views, and/or reduce her experienced guilt level when she allocates money according to this (self-serving) criterion.

While our model illustrates the broad idea behind self-serving bias, it leaves open many details about how exactly norms and beliefs map into behavior. Thus, our main contribution is in the empirical quantification of the relative importance of different channels underlying self-serving biases. Further research can use these findings to model different psychological mechanisms in more detail.

3. Design

In this paper, we report the results of two experiments. Each experiment happened over 2 days: on Day 1, participants completed real effort tasks to generate a surplus, and on Day 2, participants in the role of dictators divided the surplus. Fig. 1 displays the timeline shared by the two experiments.

For Experiment 1, we recruited 200 dictators and 300 recipients from Prolific.co. The data was collected between the 13th and 19th of July, 2020. For Experiment 2, we recruited 400 dictators and 600 recipients from Prolific.co.[2] The data was collected between the 23rd and 30th of November, 2020. These sample sizes of 100 participants per treatment were preregistered (see preregistrations at the following links: Experiment 1, Experiment 2) and larger than those of similar studies (Cappelen et al., 2007; Konow, 2000; Rodriguez-Lara & Moreno-Garrido, 2012). The final sample of 600 dictators has 43% Women and the average age of dictators is 25.24 (standard deviation 7.27). Across both experiments, we paid a completion fee of £2.85 for Day 1 and £6.15 for Day 2 plus an average bonus of around £3 per participant.

3.1. Day 1: Surplus generation

On Day 1, participants completed 8 real effort tasks. There were 4 different types of tasks: moving sliders to a predetermined position, logic questions, counting the number of zeros in a table, and solving Raven's matrices (Abeler et al., 2011; Gill & Prowse, 2012; Raven & Court, 1998). Each type of task was repeated twice. In every task, each correct answer earned a monetary reward. When completing the tasks, the participants did not know the exact monetary reward they would receive. However, they knew that they would be randomly assigned a high or low pay rate per correct answer, the amount of both pay rates, and that they would learn which pay rate applied to them at a later stage. The high pay rate was always 3 times the low pay rate, but pay rates were calibrated (based on pilot data) according to task type to result in an average surplus of £3.5 per task.

Similarly, the participants were aware that the high or low pay rate assignment would apply to all of their tasks. We checked the participants' understanding of the randomness and persistence of the pay rates with two comprehension questions that they had to answer correctly to continue with the experiment. Participants were also informed that they would be paired with other participants and that their earnings would go into a single common account, but they did not know how this would be divided.

[2] We had 16 additional Dictators that started the second day of the experiment but did not complete it. Of those 6 are Advantaged and 10 are Disadvantaged; a Fisher's exact test does not reveal a statistically significant difference in the probability of completing the experiment for these two groups ($p = 0.45$).

We recruited more recipients than dictators because dictators split the amount generated by two recipients in the Impartial trials.

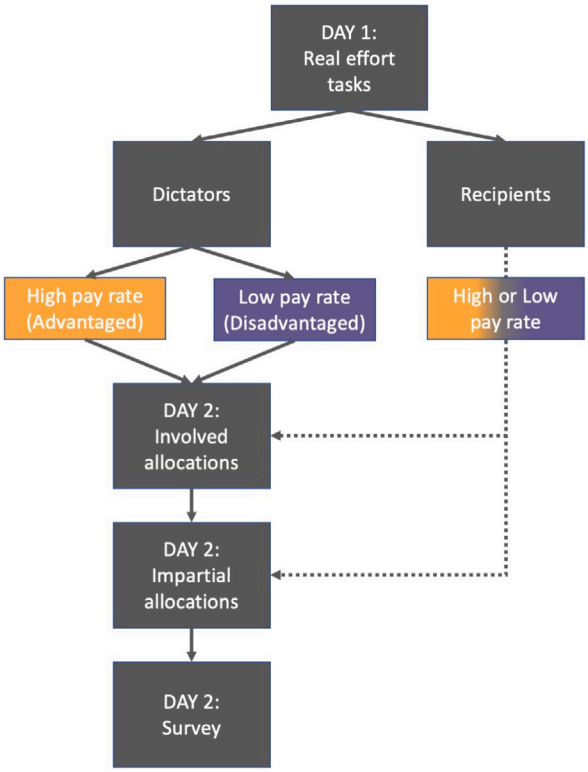

Fig. 1. Timeline for Day 1 and 2 for Both Experiments.

3.2. Day 2: Surplus division

After the Day 1 surplus generation, we split participants into dictator and recipient roles. Only the dictators were invited to Day 2, which started one day after Day 1. Day 2 was divided into 3 parts. In Part 1, dictators split earnings between themselves and recipients, termed "Involved" allocations. In Part 2, they divided the earnings between pairs of recipients, termed "Impartial" allocations. In Part 3, they answered questions about their strategies, beliefs, and perceptions of norms.

At the beginning of Day 2, dictators learned their pay rate per correct answer. We call participants who received the high pay rate "Advantaged", those with the low pay rate "Disadvantaged", and we refer to this difference as the "Status" treatment. Participants then received instructions for the Involved allocation task. The joint earnings of a pair in a task were merged into a common account, and the dictator chose how to allocate this common account between themselves and the paired recipient. Over 20 trials, the dictators were matched with different recipients, with one of the 8 tasks underlying the common account in each trial. All recipients were assigned the opposite pay rate of the dictator, thus implementing inequality in the pair. During each trial, dictators received information about the relative contributions to the common account (more on that below) and made their allocation decisions.

In the next part of Day 2, dictators made Impartial allocation decisions for two recipients. Just as in the Involved allocations, the Impartial allocations always included one Advantaged and one Disadvantaged recipient. Over 20 trials, dictators chose how to divide the common account produced by pairs of different recipients. Participants always completed the Involved trials before the Impartial trials in order to test whether self-serving biases developed in Involved decisions persisted into Impartial decisions, as in Konow (2000) and to prevent the reverse effects (Dengler-Roscher et al., 2018). Such carry-over effects are relevant outside of the lab because people typically first experience their own economic status and may develop biases dependent on that status before making more abstract, impartial decisions about fairness for others. To control for purely mechanical carry-over effects in allocation, the orientation of the slider changed for half of the participants and the slider orientation is included in regressions looking at the impact of Status on allocation.

Decisions were incentivized by implementing one of each dictator's 40 decisions. The average surplus per pair of participants in each task was £6.99 in Experiment 1 and £7.10 in Experiment 2. These amounts are approximately 1.4 times the minimum hourly wage on Prolific, so the allocation decisions had reasonably high stakes. If the decision came from the Involved allocations, the

dictator received a bonus payment equal to the amount they kept for themselves, and the recipient received the amount allocated to them. If the decision came from the Impartial allocations, the dictator received £1, and each of the two recipients received what the dictator allocated them.[3]

3.2.1. Attention measurements and differences between Experiment 1 and 2

Before every decision, the dictators had 6 seconds to look at information about the way the money in the common account was generated. Both experiments were also designed to study the role of visual attention to this information, as described in the companion paper (Amasino et al., 2021). Participants could reveal information about the number of correct answers each participant in the pair completed – merit information – as well as the monetary contribution incorporating the randomly-assigned pay rate – outcome information. This feature was implemented in MouselabWEB, so participants could reveal each piece of information by hovering their mouse cursor over the relevant labeled box (Willemsen & Johnson, 2019). Experiment 1 measured naturally occurring attention patterns with no restrictions, whereas Experiment 2 had design features to manipulate attention and investigate its causal role. In Experiment 2, there were restrictions on the length of time (400 or 1600 ms per information box) that participants could reveal either the number of correct answers or monetary contributions within the total 6 seconds to look at information, pushing them to look at one of the pieces of information longer. This attention manipulation is the only difference in the allocation decisions between Experiment 1 and Experiment 2.

In this paper, we do not analyze attention. Instead, we focus on additional measurements of norms and beliefs across experiments and attention treatments. To ensure the attention treatments do not drive the results, all the regressions in this paper control for these attention treatments.[4] Moreover, all attention treatments were designed such that participants in each condition could access information about merit and luck. We further rule out that attention might be driving our results in Supplementary sections A.6 and A.7.

3.3. Perception measurement

In Part 3 of both experiments, after the Involved and Impartial allocation decisions, we asked dictators a series of questions about their strategy, their perceptions of various fairness criteria, and their beliefs about the performance of different types of participants in the real effort tasks. For most of these variables, we conducted multiple elicitations per participant, a fact that we will leverage in the analysis. The main questions were always asked in the same order, but within a type of question, we randomized the order in which fairness rules were rated (e.g. libertarian, meritocratic, or egalitarian).[5] Moreover, we elicited participants' demographics, including gender, country, political leaning, education, and income level.

3.3.1. Personal norms of fairness

One channel for the development of self-serving biases is through the perception of what is morally appropriate behavior. In particular, Advantaged dictators may want to believe that inequalities due to luck are acceptable, while Disadvantaged ones might want to believe that these inequalities are unfair. We refer to people's fairness perceptions as "personal norms".

We obtained three independent measures of dictators' personal norms. Our main measure of personal norms is participants' ratings of the moral appropriateness of dividing according to three fairness criteria that are commonly used in the literature (Cappelen et al., 2007; Konow, 2000; Rodriguez-Lara & Moreno-Garrido, 2012): egalitarian (equal split), meritocratic (proportional to the share of correct answers), and libertarian (proportional to the share of total surplus produced - i.e. including randomly determined pay-rates).

Second, in Experiment 2 only, we asked participants to rate the moral appropriateness of allocating the surplus using different types of information. One question asked about the appropriateness of exclusively using the information about the number of correct answers, and the other about the appropriateness of exclusively using the information about the monetary contributions. While the framing is slightly different, the ratings from these questions map directly onto the appropriateness of different fairness norms. Specifically, using only information about the number of correct answers results in an allocation consistent with the meritocratic criterion, whereas using only the information about the monetary contributions results in a split consistent with the libertarian criterion.

Finally, we asked dictators an open-ended question about how they redistributed the money. Unaware of the research question, a research assistant classified whether a participant's answer referred to the egalitarian, meritocratic, or libertarian criteria. Below, we exploit the common variation in these different elicitations to address errors in the measurement of personal norms.

[3] We pre-assigned which type of trial (Involved or Impartial) would be relevant for payment, and which recipients would get the bonus to ensure that all dictators and recipients were paid a bonus based on a single allocation decision. Recipients could appear in multiple different dictators' allocation decisions.

[4] To additionally test the effect of attention, we examined the interactions between our attention treatments and norm measurements. We do not find strong interactions, so the impacts of Status on norm endorsement do not seem to be primarily driven by attention. We find that, in the merit focus treatment, Advantaged dictators are more likely to endorse personal meritocratic norms. In contrast, Disadvantaged dictators predict higher social endorsement of libertarian norms, a somewhat counterintuitive result.

[5] In Experiment 2, which has some additional elicitations compared to Experiment 1, the order of elicitations was as follows: we first asked beliefs about performance. Next, we asked about personal norms, first an open-ended question about criteria for division followed by specific questions about the appropriateness of each fairness criteria. After personal norms, we asked about overall social norms, followed by eliciting social norms specific to Advantaged or Disadvantaged dictators. Finally we asked another version of the personal norms questions about using only correct answers (merit) vs. only monetary contribution (outcome) to divide joint earnings.

3.3.2. Social norms of fairness

To understand whether participants believed that their personal norms were commonly shared, we elicited their perceptions of social norms of appropriateness related to these criteria. To do so, we used the incentivized method from Krupka and Weber (2013): participants could win a £1 reward by correctly predicting the modal response to the appropriateness question for each of the three fairness criteria.

As a further measure of social norms, we also asked participants to predict the modal answers separately for Advantaged and Disadvantaged dictators. These elicitations had two purposes. First, they served to elucidate whether people can anticipate self-serving status bias in others. Second, they help with measurement error in the mediation analysis of Section 4.3.

3.3.3. Beliefs about relative performance

Self-serving biases also arise via the formation of motivated beliefs about relative performance (Valero, 2021). Shifts in beliefs about the role of merit may affect how people think about inequality and which social norms are relevant to their decisions. In Experiment 2, we additionally elicited incentivized beliefs about two different perceptions of relative performance. To encourage them to think carefully about these questions, participants could earn a £1 bonus for a correct prediction for each case.

First, we asked dictators to report the number of trials in which the recipient outperformed them (i.e., the recipient had more correct answers). The number of correct answers is the prime criterion of merit in the experiment, so forming motivated beliefs about this topic could provide a powerful justification for keeping more of the surplus. Through our construction of the experiment rounds, dictators had a higher number of correct answers in exactly 50% of the rounds, so we can compare the answer to a baseline of 50% that participants observed in their allocation decisions.[6]

Second, we measured how participants evaluated the size of the advantage. Advantaged dictators could justify allocating larger amounts to themselves if they believe that the pay rate inequalities in the experiment are too small to make a difference in output. We elicited this belief by asking for the share of pairs in which Disadvantaged participants produced more output than the Advantaged participants. A higher share corresponds to belief in a smaller relative advantage for the Advantaged participants. We expected Advantaged participants to be more likely to believe that the treatment gap was small such that Disadvantaged participants contributed more on average, reflecting thoughts like: "The receivers I was matched with performed poorly despite having a fair chance to produce a big share of the pie, so I should be entitled keep a larger share".

4. Results

We first characterize the self-serving bias by investigating the effect of the Status treatment on dictator behavior. We then look at the causal effect of Status on personal norms, social norms, and beliefs. Finally, we look at the role of personal norms, social norms, and beliefs in explaining the self-serving bias.

Table 1 provides an overview of the means and standard deviations of the primary outcome variables.

4.1. Status and allocations

We investigate whether our results replicate those of Konow (2000). Fig. 2 displays the share of the surplus that dictators allocated to the Advantaged member of the pair, split by Status, and by Involved or Impartial allocation decisions.

4.1.1. Involved allocations

Focusing on the Involved allocations, a rank-sum test of the average share each dictator gave to the Advantaged members across rounds confirms that the two groups allocated significantly differently ($p < 0.001$). We confirm this result in regression analyses with standard errors clustered at the individual level and controls for subject characteristics, including gender, political orientation, and geographical background. Table 2, Column 1 provides the results of these regressions and shows that Advantaged dictators gave 10 percentage points more of the surplus to the Advantaged member (i.e., themselves) than Disadvantaged dictators gave to Advantaged recipients ($p < 0.001$), an effect that is almost as large as the standard deviation of allocation decisions for this group.

The fact that Advantaged dictators allocated more to themselves than Disadvantaged dictators allocated to Advantaged recipients is consistent with dictators simply keeping most of the surplus. Therefore, we look at the impact of the Status treatment on the share dictators kept for *themselves*. We see a very similar effect, with Advantaged dictators keeping 61.6% compared to Disadvantaged dictators keeping 51% (Table 1). This result is highly significant in both a rank-sum test ($p < 0.001$) as well as in a regression with controls (Table 2 - Column 2), and it replicates prior work on behavioral allocation biases whereby the participants randomly assigned a higher pay rate kept more for themselves (Deffains et al., 2016; Konow, 2000; Rodriguez-Lara & Moreno-Garrido, 2012).

In fact, the two ways of looking at the division are almost equivalent because the Disadvantaged dictators are very close to splitting the surplus 50–50 on average. This relatively even division accords with previous work showing that dictators respect earned income in their allocations (Cappelen et al., 2010; Cherry et al., 2002; Rodriguez-Lara & Moreno-Garrido, 2012).

[6] We matched dictators and recipients in such a way that dictators answered more questions correctly in 50% of the rounds. We did so to reduce the between-dictator variance in the production inputs for the common account. There is only 1 dictator for whom this matching was not possible and who had a higher number of correct answers only in 40% of the trials.

Table 1
Names and definitions of main variables.

Variable	Label	Definition	Advantaged	Disadv.
Involved allocation	% given to Adv.	The % of the common account allocated to the Advantaged participant in self-relevant decisions.	61.6 (10.8)	49.0 (13.4)
Self allocation	% Kept	The % of the common account kept by the dictator in self-relevant decisions.	61.6 (10.8)	51.0 (13.4)
Impartial allocation	% given to Adv.	The % of the common account allocated to the Advantaged recipient in impartial, self-irrelevant decisions.	55.8 (9.3)	52.2 (8.6)
Libertarian personal norms	perLib	Moral appropriateness rating (1–4) of the libertarian criterion: dividing according to monetary contributions (merit and luck).	2.75 (0.95)	2.54 (0.93)
Meritocratic personal norms	perMer	Moral appropriateness rating (1–4) of the meritocratic criterion: dividing according to the number of correct answers (merit only).	3.19 (0.82)	3.27 (0.81)
Egalitarian personal norms	perEga	Moral appropriateness rating (1–4) of the egalitarian norms: dividing evenly (regardless of merit or luck).	2.48 (0.86)	2.66 (0.88)
Libertarian social norms	socLib	Social appropriateness rating (1–4) of the libertarian criterion: dividing according to monetary contributions (merit and luck).	2.90 (0.96)	2.74 (0.95)
Meritocratic social norms	socMer	Social appropriateness rating (1–4) of the meritocratic criterion: dividing according to the number of correct answers (merit only).	3.23 (0.77)	3.29 (0.77)
Egalitarian social norms	socEga	Social appropriateness rating (1–4) of the egalitarian criterion: dividing evenly (regardless of merit or luck).	2.58 (0.86)	2.63 (0.88)
Recipient outperforming	# RecOutperf	Beliefs about the # of involved rounds (out of 20) experienced by the dictator in which the recipient had more correct answers.	8.10 (3.11)	9.68 (3.55)
Disadvantaged outcontributing	% DisOutcont	Beliefs about the % of rounds in which any Disadvantaged participant had a higher monetary contribution than an Advantaged participant.	20.95 (18.91)	25.06 (22.04)

Advantaged and Disadvantaged columns show the mean and standard deviation for each variable from both Experiment 1 and 2 for allocations and norms and from Experiment 2 only for beliefs.

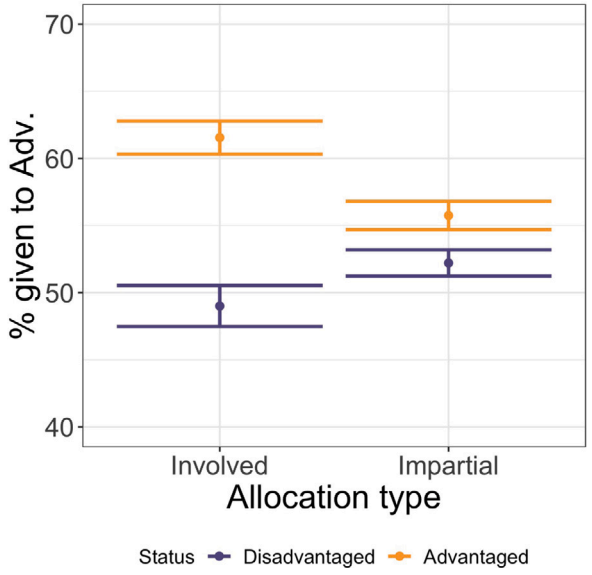

Fig. 2. Allocation by Status treatment.
The average allocations to the Advantaged member by Status (Advantaged or Disadvantaged), in both Involved decisions (left) and Impartial decisions (right). The error bars represent 95% confidence intervals based on participant-level data aggregated across trials.

Table 2
Effect of Status on allocation.

	(1) % given to Adv.	(2) % Kept	(3) % given to Adv.
Advantaged	10.0***	10.4***	3.44***
	(0.99)	(1.02)	(0.70)
Observations	11930	11930	11923
Trial type	Involved	Involved	Impartial

All models are linear regressions with standard errors clustered at the individual level. Data from Experiments 1 and 2: Involved trials in Columns (1) and (2); Impartial trials in Column (3). Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair (Columns (1) and (3)), and percentage of the surplus that the dictator kept for him/herself (Column (2)). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories), attention treatment (3 categories), task type (4 categories), slider orientation (2 categories).

4.1.2. Impartial allocations

In the Impartial allocation task, we removed the self-interest of the dictators. Any remaining favoritism by the Advantaged dictators toward Advantaged recipients thus measures the persistence of a self-serving bias that self-interest cannot explain. The right part of Fig. 2 shows evidence for such a bias, as allocation differences persist into the Impartial trials, with the Advantaged dictators still giving significantly more to Advantaged members of the pair ($p < 0.001$, rank-sum test). Column 3 of Table 2 shows that Advantaged dictators gave 3.4 percentage points more of the surplus to the Advantaged member after controlling for individual characteristics.

While statistically significant, these differences in Impartial allocations are about one-third of the difference in the Involved trials. Konow (2000) attributes these remaining differences in impartial divisions to shifting norms of fairness, an explanation we investigate in more detail below.

Result 1. *Advantaged dictators gave a larger share of the common account to themselves than Disadvantaged dictators gave to Advantaged recipients or themselves. These differences in allocations persist for Impartial choices, although the effect is less than half the size.*

4.2. Status, norms and beliefs

In this section, we investigate whether dictator Status shifted dictators' beliefs and attitudes related to allocations. In particular, we examine three sets of outcome variables: personal fairness norms, social norms, and beliefs about relative performance.

4.2.1. Personal and social norms

We first look at both personal fairness norms and anticipated social norms about the appropriateness of different fairness criteria. Fig. 3 shows the effect of Status on norm endorsement, and Table 3 gives the results of ordered logit regressions with the discrete appropriateness rating as the dependent variable (see Supplementary section A.1 for the effect of Status on our secondary norm elicitations). We find that Advantaged dictators rated libertarian norms as more appropriate on average. A rank-sum test shows the distribution of endorsement is significantly different for both personal norms ($p = 0.0044$) and social norms ($p = 0.026$). The difference in personal norms is confirmed in regressions (Table 3, Column 1), but the effect on social norms is smaller and insignificant. In addition, we find that Advantaged dictators were less likely to endorse egalitarian norms, but with statistical significance only for the personal norms elicitation (rank-sum test, $p = 0.015$), a result confirmed in our regression analyses (Table 3, Column 1). We find no statistical differences for meritocratic norms. The difference between personal and social norms indicates that subjects had some understanding that their own appropriateness ratings were biased, a finding we explore further below.

4.2.2. Can participants predict the effect of Status on norms?

We investigate whether participants were aware of the effects of Status on personal norms by asking them to predict social norms separately for Advantaged and Disadvantaged dictators. One hypothesis is that those who fail to see both perspectives and thus do not acknowledge a status bias may show stronger self-serving biases, whereas those who are aware of bias in others may reflect more and exhibit less bias (Babcock & Loewenstein, 1997). Furthermore, an awareness of how Status impacts personal norms might make people more open to interventions that attempt to reduce bias, at least for others.

To test whether participants accurately predicted the status gap in personal norms, we compare personal norms and predicted social norms in Fig. 4. We find that participants, regardless of Status, correctly anticipated that Advantaged dictators endorsed libertarian norms more highly (rank-sum tests $p < 0.001$) and Disadvantaged endorsed egalitarian norms more highly (rank-sum test $p < 0.001$). This finding suggests that participants know how Status can bias fairness views. In fact, as seen in Fig. 4, they overestimated the status biases in social norms compared to the actual differences observed in personal norms and further predicted that the Disadvantaged would be more likely to endorse meritocratic norms (rank-sum $p < 0.001$), suggesting that they anticipated others to have stronger status biases than themselves.

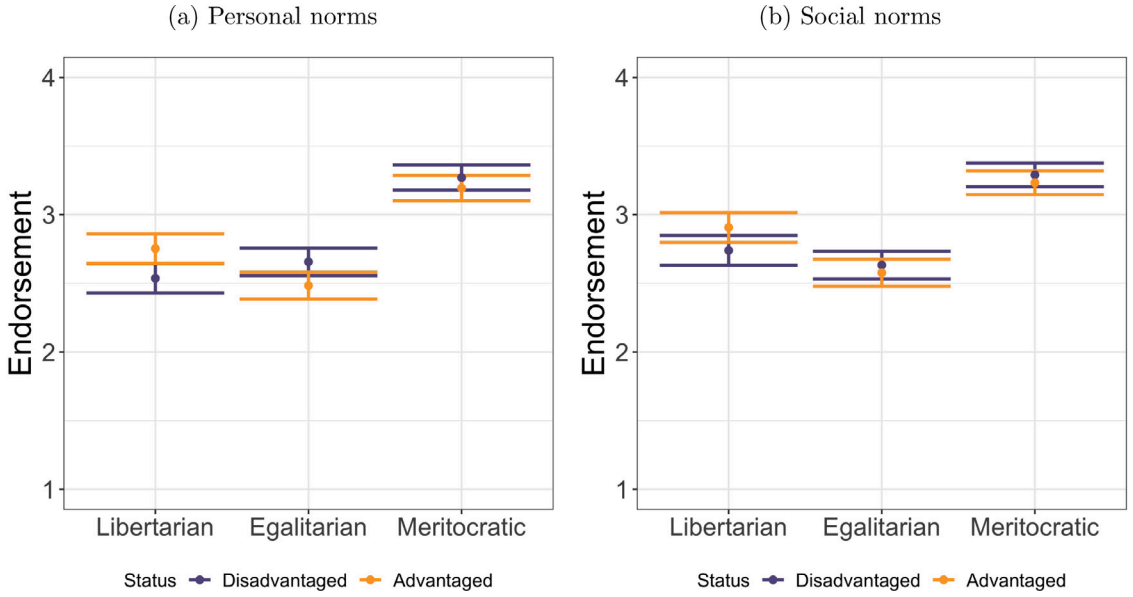

Fig. 3. Personal and social norms split by Status.
The average endorsement of libertarian, egalitarian, and meritocratic norms, both personal (left panel) and social (right panel) split by Advantaged or Disadvantaged Status. Endorsement is measured on a 1–4 scale, with 1 being very morally inappropriate and 4 being very morally appropriate. Libertarian norms mean dividing according to outcomes, including merit and luck. In contrast, egalitarian norms mean splitting evenly regardless of merit or luck. Finally, meritocratic norms mean dividing according to merit alone. Personal norms are those that participants endorse for themselves, whereas social norms are those that they predict others will endorse. The error bars represent 95% confidence intervals.

Table 3
Effect of Status on norms.

	Personal norms (1) All data	Social norms (2) All data
Panel A: Libertarian		
Advantaged	0.39*	0.29
	(0.16)	(0.16)
Panel B: Meritocratic		
Advantaged	−0.20	−0.17
	(0.16)	(0.16)
Panel C: Egalitarian		
Advantaged	−0.40*	−0.16
	(0.16)	(0.15)
Observations	600	600

All models are ordered logits. Data from Experiment 1 and Experiment 2. Dependent variable: social or moral acceptability of a norm (1 very inappropriate, 2 somewhat inappropriate, 3 somewhat appropriate, 4 very appropriate). Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories), attention treatment (3 categories).

Despite the awareness of how Status influenced self-serving biases in allocations, we do not find any relationship between predicting larger status gaps in norms and allocation choices (Spearman's correlations between the predicted gap and the % allocated to the Advantaged in Impartial decisions: libertarian gap: $\rho = -0.07$, $p = 0.07$; meritocratic gap: $\rho = -0.005$, $p = 0.91$; egalitarian gap: $\rho = 0.02$, $p = 0.67$). This lack of relationship between predicted status gaps in social norms and allocations suggests that participants may be subject to similar self-serving biases in allocations that they predict in others. Nevertheless, the awareness that Status impacts fairness norms may matter – despite the lack of correlation with one's own bias – as it could lead to acceptance of interventions to reduce bias in "others", even if people think that they are uniquely immune to such biases.

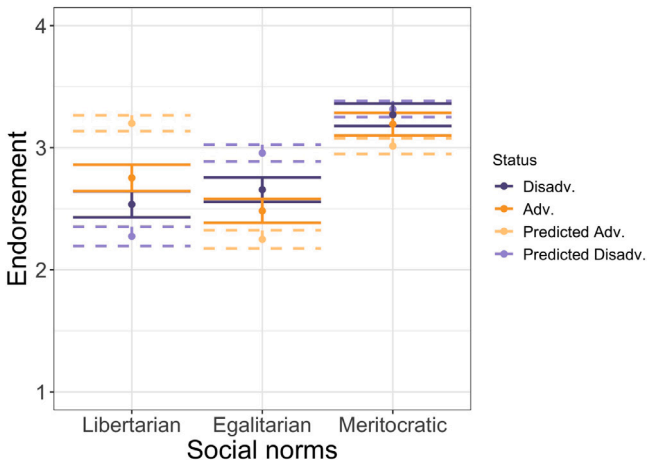

Fig. 4. Predicted social norms vs. actual personal norms split by Status.
The average endorsement of libertarian, egalitarian, and meritocratic norms. Predicted social norms for Advantaged vs. Disadvantaged and actual personal norms are displayed. Endorsement is measured on a 1-4 scale with 1 being very morally inappropriate and 4 being very morally appropriate. The error bars represent 95% confidence intervals. Note: one's own Status has minimal effect on the predictions of social norms by Status, so the predictions have been collapsed across participants' Status.

4.2.3. Beliefs about relative performance

We now turn to beliefs about relative performance, elicited only in Experiment 2. Fig. 5 shows an overview of the mean beliefs and confidence intervals across Status treatments. All participants showed some bias toward underestimating the number of rounds in which the recipients outperformed them (the true answer was 50% for all participants except 1 for whom it was 40%), but this is particularly pronounced in Advantaged dictators. In line with the tendency to form motivated beliefs ~ 80% of Advantaged dictators indicated that the other participant did equally well or worse than them, compared to only 60% of Disadvantaged dictators (rank-sum test $p < 0.001$). This difference in beliefs is confirmed by an OLS regression displayed in Table 4, Column 1. It shows that Advantaged dictators believed that recipients answered more questions correctly in 1.6 fewer rounds (about 8% of the total rounds) than Disadvantaged dictators did ($p < 0.001$). Because beliefs about performance were asked after the allocation decisions, participants could simply have remembered their performance on each round to answer this question, which may reduce the bias in beliefs compared to studies with more ambiguity (Deffains et al., 2016; Valero, 2021). Nevertheless, the difference in beliefs depending on the Advantaged Status suggests that biased beliefs (or memories) are still present to some extent even with full information, as was also found in Espinosa et al. (2020).

We then look at beliefs about the size of the disadvantage, as measured bys beliefs about the probability that a Disadvantaged member would out-contribute an Advantaged member. Both groups of subjects overestimated this probability: the real chance was 6.8% while they believed it to be 23% (t-test, $p < 0.001$). In addition, we do not find support for the idea that Advantaged dictators underestimated their random advantage more than Disadvantaged dictators did to downplay the role of luck (rank-sum test $p = 0.068$). An OLS regression analysis (Table 4, Column 2) finds that the beliefs go in the opposite direction of what would be expected: Advantaged dictators thought it less likely (by about 4 percentage points) that Disadvantaged dictators outperformed Advantaged dictators in terms of monetary contributions ($p = 0.05$). A plausible alternative explanation is that this result represents a different form of self-serving bias, whereby Advantaged dictators interpreted larger monetary contributions as signifying a higher deservingness instead of a larger artificial advantage.[7]

Result 2. *Advantaged dictators personally endorsed egalitarian sharing rules less and libertarian sharing rules more than Disadvantaged dictators. Overall social norm perceptions were similar, but statistically non-significant; however, for social norms split by Dis(Advantage), participants predicted significant status biases regardless of their own Status. Advantaged dictators were also more likely to believe that they outperformed and out-contributed in the task.*

[7] Yet another explanation is that Advantaged dictators were so convinced of being better at the task that their beliefs about the size of advantage did not reverse this. We can check this interpretation in the same model shown in Column 2 by controlling for the dictator's beliefs about the number of rounds in which the Disadvantaged member of the pair answered more questions correctly than the Advantaged member. This additional control does not change the results from Column 2, indicating that the beliefs about the recipients' correct answers are not driving the result.

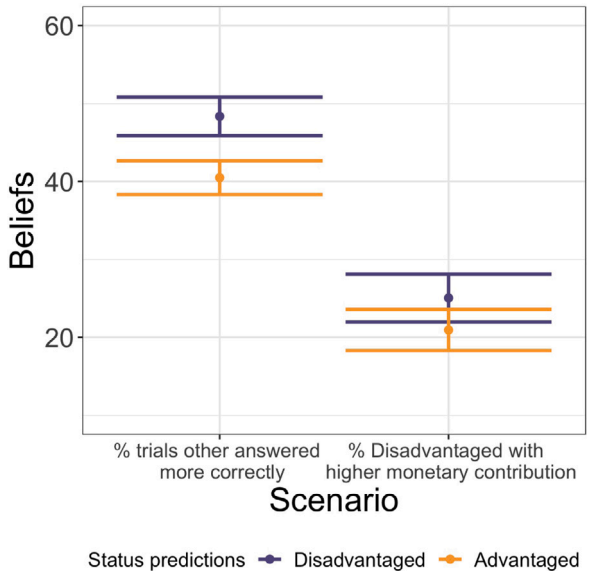

Fig. 5. Beliefs about performance by Status.
The average beliefs about performance split by Status (Advantaged or Disadvantaged). On the left, the participants state their beliefs about the % of Involved trials on which the recipients answered more questions correctly than them. In contrast, on the right, the participants estimate the % of trials on which Disadvantaged participants had a higher monetary contribution than Advantaged participants. The error bars represent 95% confidence intervals.

Table 4
Effect of Status on beliefs.

	(1) # RecOutperf	(2) % DisOutcont
Advantaged	−1.60***	−4.12*
	(0.35)	(2.09)
Observations	400	400

All models are linear regressions. Data from Experiment 2. Dependent variables: (1) # RecOutperf: the dictators beliefs about the number of rounds in which the recipient answered more questions correctly of the dictator him/herself. (2) % DisOutcont: Beliefs about the % chance that any Disadvantaged dictator contributed a higher monetary contribution than an Advantaged dictator on any round. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories), attention treatment (3 categories).

4.3. Do norms and beliefs explain allocations?

We now try to quantify how much of the self-serving bias can be explained by variations in norms and beliefs. As our measure of self-serving bias, we use the effect of Status on allocations in Impartial trials. Since self-interest has been eliminated as a motive in these trials, this status effect is most likely to reflect internalized shifts in fairness or beliefs. We perform this analysis using the common Difference of Coefficients Approach for mediation analysis (Judd & Kenny, 1981).

4.3.1. Correcting for measurement error

A key problem in mediation analyses comes from measurement error. Gillen et al. (2019) show that small noise in the measurement of the mediators – in our case norms and beliefs – can lead to a severe underestimation of their mediating role. To address this problem, we leverage our multiple elicitations in Experiment 2, where we have repeated elicitations of both personal and social norms.

For personal and social norms, we exploit the Instrumental Variables (IV) approach suggested by Gillen et al. (2019) to isolate the common variation in the multiple elicitations. More precisely, we instrument our measures of personal norms with the alternative elicitations of the appropriateness of using different types of information to divide the common account, which measure the same

Table 5
Impartial allocations to Advantaged recipients controlling for norms and beliefs.

	(1) % given to Adv.	(2) % given to Adv.	(3) % given to Adv.	(4) % given to Adv.
Advantaged	3.04***	1.76	2.71**	2.82**
	(0.83)	(1.04)	(0.85)	(0.93)
Personal norms		✓		
		(37.07)		
Social norms			✓	
			(10.38)	
Beliefs				✓
				(2.27)
F-statistic		10.4	21.0	
Observations	400	400	400	400

Linear and 2SLS regressions with standard errors clustered at the individual level. Data: Impartial trials from Experiment 2. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. Columns (1) and (4) are linear regressions, Columns (2) and (3) are 2SLS models. In Column (2), the instrumented variables are perLib, perEga, and perMer; the instruments are our alternative personal norms elicitations. In Column (3), the instrumented variables are socLib, socMer, and socEga; the instruments are participants' perceptions of the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. Column (4) contains two beliefs variables. The first is "% DisOutcon". The second is "# DisOutperfom", which indicates the dictators' beliefs about the number of rounds in which the disadvantaged member of the pair answered more questions correctly in the task. This variable is generated from a simple transformation of the variable "# RecOutperf". The variables mentioned in this caption are defined in Table 1. List of controls common to all regressions: percentage of the joint number of correct answers due to the Advantaged recipient, age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories), attention treatment (2 categories), slider orientation (2 categories). Standard errors in parentheses under the coefficients for "Advantaged"; statistics for the joint significance of norms $\chi^2(3)$ or beliefs $\chi^2(2)$ shown in parentheses below the check marks. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The F-statistic is the Kleibergen–Paap rk Wald F statistic.

underlying construct as discussed in Section 3. As an additional instrument, we use our coding of the open-ended answers about the use of fairness criteria. To instrument the social norms, we used the participants' predictions about the social norms of (a) the Advantaged dictators and (b) the Disadvantaged dictators. These are valid instruments as they jointly provide information on the perceptions of social norms endorsed in the universe of dictators.

For beliefs, we cannot follow the same approach: our two belief elicitations concern related but not completely overlapping aspects of performance. Hence, the best we can do is to enter both our beliefs measures linearly as controls. As Gillen et al. (2019) show, this approach should still reduce concerns due to measurement errors, but it leaves more room for error.

Supplementary section A.4 quantifies the importance of correcting for measurement error in our setting by comparing estimates with and without corrections. It shows that we would underestimate the explanatory power of personal norms by a factor of 2.5 without correction for measurement error.

4.3.2. Mediation results

Table 5 displays the results of linear regressions using the data from Experiment 2, in which we have multiple elicitations of both personal and social norms. This analysis uses observations averaged at the individual level because the variability in norms and beliefs is only between and not within subjects. We first establish our baseline result without controlling for norms or beliefs. Column 1 shows that the overall bias in Experiment 2 is just over 3 percentage points. Note that this is similar to the estimate of 3.4 over both experiments presented in Table 2, indicating that the exclusion of Experiment 1 does not change the results.

We next examine whether personal norms, social norms, or beliefs explain the self-serving bias in allocations by Status. Column 2 displays a two stages least square estimation that includes personal norms. In the first stage, the F-statistic is 10.4, indicating that our instruments are highly relevant. To compare the change in coefficient on Advantaged between stages, we use the method for comparing coefficients of nested models described in Clogg et al. (1995). We find that in the second stage, the coefficient for Advantaged drops to 1.76, is no longer significantly different from zero ($p = 0.090$), and is significantly different from the 3.04 coefficient in Column (1) ($t(372) = 2.83$, $p = 0.005$). Column 3 displays the same analysis as in Column 2 but with social rather than personal norms. In the first stage, the F-statistic is 21.0, indicating, once again, a small expected bias in the estimates. The coefficient for Advantaged decreases to 2.71, remains both significantly different from zero ($p = 0.001$) and not significantly different from the coefficient in Column 1 ($t(372) = 1.25$, $p = 0.21$). Finally, Column 4 displays a linear regression that controls for beliefs about performance. The coefficient for Advantaged becomes 2.82, remains significantly different from zero ($p = 0.003$) and not significantly different from the coefficient in Column 1 ($t(372) = 0.53$, $p = 0.60$).

We can judge the explanatory power of the three psychological variables by comparing the coefficients for being Advantaged in the different columns. Personal norms explain 42% of the self-serving bias, social norms explain about 11%, and beliefs explain about 6%. To compute these numbers, we use Table 5, and we take the difference between the coefficient for Advantaged in Column (1) and the coefficient for Advantaged in the column where we control for a given psychological channel. We then divide the result for the coefficient for Advantaged in Column (1). For example, the effect of Status that passes via personal norms is given by

$(3.04 - 1.76)/3.04 = 0.42$. Where 3.04 is the total effect of being Advantaged on allocation from Column (1); 1.76 is the effect of Status on the allocation that does not pass via personal norms, from Column (2). The limited explanatory power of social norms is in line with the weak effect of Status on these norms. Instead, the explanatory power of beliefs might be underestimated, as we cannot entirely eliminate measurement bias from the belief variables.

4.3.3. Do norms and beliefs predict allocations?

Next to the mediation, we look at the direct connection between norms, beliefs, and allocations. We do so by looking at the coefficients for these variables in the regressions described in Table 5, the same we used for the mediation analysis above. To reduce the risk of false positives, we use 3 joint statistical tests rather than testing the significance of each of the 8 coefficients separately and therefore do not display the individual coefficients in Table 5 (these coefficients are reported in Supplementary Table 6). Moreover, we compare the p-values with the Bonferroni adjusted significance level of $\alpha = 0.0167$, obtained by dividing the canonical $\alpha = 0.05$ by 3, the number of tests we are running. The test of the joint significance of personal norms in Column (2) rejects the null hypothesis that none of the three personal norms correlates with allocation decisions ($\chi^2(3) = 37.07$, $p < 0.001$). The test on the joint significance of social norms in Column (3) rejects a similar null hypothesis for social norms ($\chi^2(3) = 10.38$, $p = 0.0156$), showing that social norms correlate with decisions as well. Finally, the test in Column (4) fails to reject the null hypothesis that neither of the two beliefs correlates with behavior ($\chi^2(2) = 2.27$, $p = 0.32$). Supplementary Table 7 shows evidence that the effect of personal norms on behavior depends on the dictator's Status.

4.3.4. Robustness

Table 5 does not include specifications that combine norms and beliefs to estimate the total amount of self-serving biases that are mediated by these variables. The reason is that instrumenting the personal and the social norms at the same time results in a first-stage F-statistic below 2 and, hence, in a large expected bias in the estimates, likely due to collinearity between some of the instruments. An additional limitation is that the analysis assumes a linear relationship between norms, beliefs, and redistribution. To address these limitations, the regressions reported in Supplementary Table 4 control for beliefs entering them as 4th-degree polynomials and for personal and social norms entering them as dummy variables for each possible norm rating. Moreover, to limit the bias due to measurement error, every set of elicitations available for the norms is included in the regression (Gillen et al., 2019). The results of this alternative mediation analysis are similar to those from Table 5. Moreover, this analysis shows that our ability to explain self-serving biases does not change much if we control for personal norms, social norms and beliefs jointly.

Result 3. *Shifts in personal norms capture 42% of the self-serving bias in impartial decisions; social norms capture 11%, whereas self-serving beliefs about performance capture at least 6%.*

5. Discussion

A number of design features of our study have implications for the generalizability and interpretation of our findings, which we discuss in detail below.

First, we always elicit all of our constructs after participants make their allocation choices. This means that participants may want to justify their allocations when answering these constructs. In the literature on the order effects of choice and elicitations, d'Adda et al. (2016) find that behavior may shift if norms are elicited first but that norms (especially incentivized social norms) are less likely to change regardless of whether they are elicited before or after behavior, suggesting that our choice to elicit norms after allocation choices is the cleanest design. However, Rustichini and Villeval (2014) find a more a bi-directional relationship between personal norm elicitations and choice where norms elicited after choice may be used to justify decisions. A recent paper by Charness et al. (2021) on eliciting beliefs also discusses mixed evidence on whether elicitations bias subsequent choices, arguing that more research is needed and that, in light of the ambiguity, the more important metric (choice or belief) should be elicited first. We consider allocation choices as our primary measure and thus we elicited them first. Similarly, we elicit beliefs, personal norms, and social norms in the same order across all participants. We chose the order to try to minimize spillover effects by putting questions that might be perceived as more leading later in the order of elicitation. Nevertheless, there could be some influence of earlier questions on later ones that is not fully accounted for in our analyses.

Second, a few features of our design may push toward personal norms explaining the most variance. One feature is the anonymous setting which eliminates social stakes such as reputation or punishment, giving personal norms the best chance of influencing behavior (Eckel et al., 2022; Fehr & Gächter, 2000; Salazar et al., 2022). The anonymity or visibility of context, in addition to other situational factors like the salience of personal or social norms, likely affects the extent to which people feel obligated to follow social norms (Cialdini et al., 1991; Kallgren et al., 2000). The lack of anonymity may be why Bursztyn et al. (2020) and Sparkman et al. (2022) find that social norms dominate personal norms in settings where the fear of sanctions may reduce the relative importance of personal norms. Ajzen and Fishbein (1970) also show that even the simple framing of a prisoner's dilemma as competitive or cooperative can impact the relative weight of social norms vs. personal norms. That said, more social exposure could also reduce bias in social norms, as the subjects would have a bigger incentive to correctly anticipate the reactions of others to their choices. Given that we were primarily interested in the justifications people use even when there are no consequences, because decisions like voting about redistribution are typically made in private, we designed the study to focus on self-image in an anonymous setting. Another design feature potentially making personal norms more important is that personal norms are not incentivized, allowing

them to be more subject to a consistency bias and justification since there is little cost to manipulating them, as opposed to social norms, which are incentivized.

Finally, the participants make the involved allocations before the impartial ones, which may strengthen the biased allocations in the impartial decisions. We are interested in how developing self-serving fairness principles in the involved decisions spills over into impartial decisions due to cognitive dissonance, hence the choice of the ordering. However, putting the impartial allocation before the involved could show whether cognitive dissonance works in the other direction, whereby participants stick to more impartial fairness rules even in the involved decisions (or shift their fairness rule from the beginning, anticipating the effect on involved decisions). In the literature, Dengler-Roscher et al. (2018) directly test how the order of involved vs. impartial decisions affects allocations after a real-effort task, albeit in a situation without luck. They find larger deviations from meritocratic divisions for involved allocations as compared with impartial allocations and less deviation from these meritocratic divisions when impartial allocations are made first (but only for participants who do not have prior experience of allocation tasks). Further, Valero (2021) shows that knowing there will be a later opportunity to redistribute does not shift beliefs about the underlying luck or merit of success, suggesting that people are not strategic enough to manipulate their beliefs when they could financially benefit from it later. On the contrary, Saccardo and Serra-Garcia (2023) show that the formation of an unbiased opinion reduces subsequent corruptibility of experimental subjects when giving financial advice, but they also find that subjects anticipate such effects. Together, these findings suggest that putting impartial allocations first might reduce the effect of status on norms and self-serving allocation biases, an avenue for further research.

6. Conclusion

In this paper, we investigate the role of norms and beliefs in explaining self-serving biases. We find evidence that randomly advantaged participants are less likely to believe that redressing inequalities due to luck is morally appropriate and more likely to overestimate their economic performance. However, the random advantage leads to smaller and insignificant shifts of social norms. Variation in norms and beliefs explains around 42% of the self-serving bias in allocation behavior, primarily driven by the impact of personal norms. Our design allows precise quantitative estimates thanks to a reduction in measurement error, which more than doubles the impact of such norms relative to uncorrected estimates.

These results show that economic status has an effect on personal norms as well as beliefs, with shifts in personal norms of fairness emerging as the most important explanation of self-serving biases. This suggests that modeling efforts should focus on this particular psychological mechanism. So far, there does not seem to be consensus as to how to best incorporate such norms in economic models. Policy-makers who aim to reduce self-serving biases about redistribution should also focus on personal norms, for instance through moral persuasion campaigns that have been successful in reducing ethnicity-based biases (Blouin & Mukand, 2019). While rewiring people's conceptions of what constitutes socially acceptable behavior might be difficult to accomplish and less impactful, our findings suggest that campaigns can be effective by targeting personal norms, which are relatively elastic.

While personal norms can explain a large part of self-serving biases, almost 60% of the bias in our experiment remains unexplained. One additional factor not explored in this study is the potential for in-group favoritism, which has been found to play an important role in differential allocations in prior studies (Cassar & Klein, 2019; Dorin et al., 2021). Future work may further reduce measurement bias in beliefs and account for other mechanisms, like in-group favoritism, that this study does not explore. Moreover, this paper performs a correlational mediation analysis that does not allow for causal claims on the relationship between the mediators – norms and beliefs – and behavior (Imai et al., 2013). Future work might study these relationships more directly, by developing experimental designs that manipulate beliefs or norms.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data and analysis code can be found at: https://osf.io/qjy6u/files/osfstorage.

Acknowledgments

We thank Vinska Talita Johan, Antonia Kurz, and Cristina Figueroa for excellent research assistance, Jan Hausfeld, Kosuke Imai, Margarita Leib, Caroline Liqui Lung, Joep Sonnemans, Ivan Soraperra, Jan Stoop, and Egon Tripodi as well as seminar participants at the University of Amsterdam and the University of Pittsburgh, the ViProc conference, and the BBU flash talks workshop for very useful comments. Joël van der Weele gratefully acknowledges funding by the NWO, Netherlands in the context of VIDI grant 452-17-004. Davide Domenico Pace gratefully acknowledges financial support by Deutsche Forschungsgemeinschaft, Germany through CRC TRR 190 (project number 280092119). The data and analysis code for this project can be found at this link https://osf.io/qjy6u/files/osfstorage

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.joep.2023.102654.

References

Abeler, Johannes, Falk, Armin, Goette, Lorenz, & Huffman, David (2011). Reference points and effort provision. *American Economic Review*, *101*(2), 470–492. http://dx.doi.org/10.1257/aer.101.2.470.

Ajzen, Icek, & Fishbein, Martin (1970). The prediction of behavior from attitudinal and normative variables. *Journal of Experimental Social Psychology*, *6*(4), 466–487.

Almås, Ingvild, Cappelen, Alexander W, Sørensen, Erik Ø., & Tungodden, Bertil (2022). Global evidence on the selfish rich inequality hypothesis. *Proceedings of the National Academy of Sciences*, *119*(3), https://doi.org/10.1073/pnas.2109690119.

Amasino, Dianna, Pace, Davide, & van der Weele, Joël J. (2021). *Fair shares and selective attention*: *Tech. Rep.*, (2021–066), Tinbergen Discussion Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3890037.

Andre, Peter, Boneva, Teodora, Chopra, Felix, & Falk, Armin (2021). *Fighting climate change: the role of norms, preferences, and moral values*: *Tech. Rep.*, (No. DP16343), CEPR Discussion Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3886831.

Babcock, Linda, & Loewenstein, George (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic perspectives*, *11*(1), 109–126.

Babcock, Linda, Loewenstein, George, Issacharoff, Samuel, & Camerer, Colin (1995). Biased Judgments of Fairness in Bargaining. *The American Economic Review*, *85*(5), 1337–1343, http://www.jstor.org/stable/2950993.

Bašić, Zvonimir, & Verrina, Eugenio (2021). *Personal norms — and not only social norms — shape economic behavior*: *Tech. Rep.*, (ID 3720539), Rochester, NY: Social Science Research Network, https://papers.ssrn.com/abstract=3720539.

Bénabou, Roland, & Tirole, Jean (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652–1678, http://www.nber.org/papers/w11535.

Bicchieri, Cristina, Dimant, Eugen, & Sonderegger, Silvia (2023). It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion. *Games and Economic Behavior*, *138*, 321–354, https://doi.org/10.1016/j.geb.2023.01.005.

Blouin, Arthur, & Mukand, Sharun W. (2019). Erasing ethnicity? propaganda, nation building, and identity in Rwanda. *Journal of Political Economy*, *127*(3), 1008–1062. http://dx.doi.org/10.1086/701441.

Bortolotti, Stefania, Soraperra, Ivan, Sutter, Matthias, & Zoller, Claudia (2017). *Too lucky to be true - fairness views under the shadow of cheating*: *Tech. Rep.*, (ID 3014734), Rochester, NY: Social Science Research Network, https://papers.ssrn.com/abstract=3014734.

Bradley, Gifford W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question.. *Journal of Personality and Social Psychology*, *36*(1), 56.

Bursztyn, Leonardo, González, Alessandra L., & Yanagizawa-Drott, David (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review*, *110*(10), 2997–3029. http://dx.doi.org/10.1257/aer.20180975.

Cappelen, Alexander W., Hole, Astri Drange, Sørensen, Erik Ø., & Tungodden, Bertil (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, *97*(3), 818–827.

Cappelen, Alexander W., Konow, James, Sørensen, Erik Ø., & Tungodden, Bertil (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, *103*(4), 1398–1413.

Cappelen, Alexander W., Moene, Karl Ove, Skjelbred, Siv-Elisabeth, & Tungodden, Bertil (2023). The merit primacy effect. *The Economic Journal*, *133*(651), 951–970, https://doi.org/10.1093/ej/ueac082.

Cappelen, Alexander W., Sørensen, Erik Ø., & Tungodden, Bertil (2010). Responsibility for what? Fairness and individual responsibility. *European Economic Review*, *54*(3), 429–441. http://dx.doi.org/10.1016/j.euroecorev.2009.08.005.

Cassar, Lea, & Klein, Arnd H. (2019). A matter of perspective: How failure shapes distributive preferences. *Management Science*, *65*(11), 5050–5064. http://dx.doi.org/10.1287/mnsc.2018.3185.

Charness, Gary, Gneezy, Uri, & Rasocha, Vlastimil (2021). Experimental methods: Eliciting beliefs. *Journal of Economic Behaviour and Organization*, *189*, 234–256, https://doi.org/10.1016/j.jebo.2021.06.032.

Cherry, Todd L., Frykblom, Peter, & Shogren, Jason F. (2002). Hardnose the dictator. *American Economic Review*, *92*(4), 1218–1221.

Cialdini, Robert B., Kallgren, Carl A., & Reno, Raymond R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology*: *Vol. 24*, (pp. 201–234). Elsevier.

Clogg, Clifford C., Petkova, Eva, & Haritou, Adamantios (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100*(5), 1261–1293. http://dx.doi.org/10.1086/230638.

Cohn, Alain, Jessen, Lasse J., Klasnja, Marko, & Smeets, Paul (2019). *Why do the rich oppose redistribution? An experiment with America's top 5%*: *Tech. Rep.*, (ID 3395213), Rochester, NY: Social Science Research Network, https://papers.ssrn.com/abstract=3395213.

d'Adda, Giovanna, Drouvelis, Michalis, & Nosenzo, Daniele (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, *62*, 1–7, https://doi.org/10.1016/j.socec.2016.02.003.

Deffains, Bruno, Espinosa, Romain, & Thöni, Christian (2016). Political self-serving bias and redistribution. *Journal of Public Economics*, *134*, 67–74. http://dx.doi.org/10.1016/j.jpubeco.2016.01.002.

Dengler-Roscher, Kathrin, Montinari, Natalia, Panganiban, Marian, Ploner, Matteo, & Werner, Benedikt (2018). On the malleability of fairness ideals: Spillover effects in partial and impartial allocation tasks. *Journal of Economic Psychology*, *65*, 60–74, https://doi.org/10.1016/j.joep.2017.11.001.

Di Tella, Rafael, Galiant, S., & Schargrodsky, E. (2007). The formation of beliefs: Evidence from the allocation of land titles to squatters. *Quarterly Journal of Economics*, *122*(1), 209–241. http://dx.doi.org/10.1162/qjec.122.1.209.

Dorin, Camille, Hainguerlot, Marine, Huber-Yahi, Hélène, Vergnaud, Jean-Christophe, & de Gardelle, Vincent (2021). How economic success shapes redistribution: The role of self-serving beliefs, in-group bias and justice principles. *Judgment and Decision Making*, *16*(4), 932–949. http://dx.doi.org/10.1017/S1930297500008032.

Durante, Ruben, Putterman, Louis, & van der Weele, Joël (2014). Preferences for redistribution and perception of fairness: An experimental study. *Journal of the European Economic Association*, *12*(4), 1059–1086. http://dx.doi.org/10.1111/jeea.12082.

Eckel, Catherine C., Fatas, Enrique, & Kass, Malcolm (2022). Sacrifice: An experiment on the political economy of extreme intergroup punishment. *Journal of Economic Psychology*, *90*, Article 102486, https://doi.org/10.1016/j.joep.2022.102486.

Espinosa, Romain, Deffains, Bruno, & Thöni, Christian (2020). Debiasing preferences over redistribution: An experiment. *Social Choice and Welfare*, *55*(4), 823–843. http://dx.doi.org/10.1007/s00355-020-01265-z.

Fehr, Ernst, & Gächter, Simon (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Gill, David, & Prowse, Victoria (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, *102*(1), 469–503. http://dx.doi.org/10.1257/aer.102.1.469.

Gillen, Ben, Snowberg, Erik, & Yariv, Leeat (2019). Experimenting with measurement error: Techniques with applications to the Caltech Cohort study. *Journal of Political Economy*, *127*(4), 1826–1863. http://dx.doi.org/10.1086/701681.

Hochleitner, Anna (2022). *Fairness in times of crisis: Negative shocks, relative income and preferences for redistribution*: *Tech. Rep.*, CeDEx Discussion Paper Series, http://hdl.handle.net/10419/261246.

Hvidberg, Kristoffer B., Kreiner, Claus, & Stantcheva, Stefanie (2020). *Social position and fairness views*: *Tech. Rep.*, National Bureau of Economic Research, http://www.nber.org/papers/w28099.

Imai, Kosuke, Tingley, Dustin, & Yamamoto, Teppei (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 5–51, https://doi.org/10.1111/j.1467-985X.2012.01032.x.

Judd, Charles M., & Kenny, David A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*(5), 602–619.

Kallgren, Carl A., Reno, Raymond R., & Cialdini, Robert B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, *26*(8), 1002–1012.

Konow, James (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, *90*(4), 1072–1092, http://www.jstor.org/stable/10.2307/117326.

Koo, Hyunjin J, Piff, Paul K, & Shariff, Azim F (2023). If I could do it, so can they: Among the rich, those with humbler origins are less sensitive to the difficulties of the poor. *Social Psychological and Personality Science*, *14*(3), 333–341. http://dx.doi.org/10.1177/19485506221098921.

Krawczyk, Michał (2010). A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics*, *94*(1), 131–141. http://dx.doi.org/10.1016/j.jpubeco.2009.10.003.

Krupka, Erin L., & Weber, Roberto A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*(3), 495–524. http://dx.doi.org/10.1111/jeea.12006.

Lefgren, Lars J., Sims, David P., & Stoddard, Olga B. (2016). Effort, luck, and voting for redistribution. *Journal of Public Economics*, *143*, 89–97. http://dx.doi.org/10.1016/j.jpubeco.2016.08.012.

Lobeck, Max (2023). *Motivating beliefs in a just world*: *Tech. Rep.*, (ID 3720539), Rochester, NY: Social Science Research Network, https://ssrn.com/abstract=4369242.

Messick, David M., & Sentis, Keith P. (1979). Fairness and preference. *Journal of Experimental Social Psychology*, *15*(4), 418–434.

Miller, Dale T., & Ross, Michael (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213.

Piketty, Thomas (2020). *Capital and ideology*. Harvard University Press.

Raven, John C, & Court, John Hugh (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press Oxford.

Rodriguez-Lara, Ismael, & Moreno-Garrido, Luis (2012). Self-interest and fairness: Self-serving choices of justice principles. *Experimental Economics*, *15*(1), 158–175, https://doi.org/10.1007/s10683-011-9295-3.

Rustichini, Aldo, & Villeval, Marie Claire (2014). Moral hypocrisy, power and social preferences. *Journal of Economic Behavior & Organization*, *107*, 10–24, https://doi.org/10.1016/j.jebo.2014.08.002.

Saccardo, Silvia, & Serra-Garcia, Marta (2023). Enabling or limiting cognitive flexibility? Evidence of demand for moral commitment. *American Economic Review*, *113*(2), 396–429. http://dx.doi.org/10.1257/aer.20201333.

Salazar, Miguel, Shaw, Daniel Joel, Czekóová, Kristína, Staněk, Rostislav, & Brázdil, Milan (2022). The role of generalised reciprocity and reciprocal tendencies in the emergence of cooperative group norms. *Journal of Economic Psychology*, *90*, Article 102520, https://doi.org/10.1016/j.joep.2022.102520.

Sandel, Michael J. (2020). *The tyranny of merit: What's become of the common good?*. Allen Lane London.

Schwardmann, Peter, Tripodi, Egon, & Van der Weele, Joël J. (2022). Self-persuasion: Evidence from field experiments at international debating competitions. *American Economic Review*, *112*(4), 1118–1146. http://dx.doi.org/10.1257/aer.20200372.

Sparkman, Gregg, Geiger, Nathan, & Weber, Elke U. (2022). Americans experience a false social reality by underestimating popular climate policy support by nearly half. *Nature Communications*, *13*(1), 4779, https://doi.org/10.1038/s41467-022-32412-y.

Suhay, Elizabeth, Klašnja, Marko, & Rivero, Gonzalo (2021). Ideology of affluence: Explanations for inequality and economic policy preferences among rich americans. *The Journal of Politics*, *83*(1), 367–380, https://doi.org/10.1086/709672.

Thøgersen, John (2008). Social norms and cooperation in real-life social dilemmas. *Journal of Economic Psychology*, *29*(4), 458–472.

Ubeda, Paloma (2014). The consistency of fairness rules: An experimental study. *Journal of Economic Psychology*, *41*, 88–100. http://dx.doi.org/10.1016/j.joep.2012.12.007.

Valero, Vanessa (2021). Redistribution and beliefs about the source of income inequality. *Experimental Economics*, *25*(3), 876–901. http://dx.doi.org/10.1007/s10683-021-09733-8.

Willemsen, Martijn C., & Johnson, Eric J. (2019). Observing cognition with MouselabWEB. In *A handbook of process tracing methods* (pp. 76–95).