Self-serving Bias in Redistribution Choices: Accounting for Beliefs and Norms

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

[a]CREED, Amsterdam School of Economics, University of Amsterdam, Amsterdam, the Netherlands

[b]Ludwig-Maximilians-Universität München

[c]Tinbergen Institute, the Netherlands

Abstract

People with higher-incomes tend to support less redistribution than lower-income people. This has been attributed not only to self-interest, but also to psychological mechanisms including 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 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 beliefs about 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.

Keywords: Redistribution; self-serving bias; fairness; norms; online experiments.
JEL codes: C91, D63, D83.

Acknowledgements: 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 Virtual Process Tracing conference, and the BBU flash talks workshop for very useful comments. Joël van der Weele gratefully acknowledges funding by the NWO in the context of VIDI grant 452-17-004.

Declaration of interest: None.

Corresponding author: Dianna R. Amasino, d.r.amasino@uva.nl, Center for Experimental Economics and Political Decision-Making (CREED), Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, Netherlands

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 (Koo et al., 2022; Suhay et al., 2020; Cohn et al., 2019; Konow, 2000; Di Tella et al., 2007). 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 (Piketty, 2020; Sandel, 2020; Babcock et al., 1995; 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., 2019), or motivated beliefs about merit and returns to effort (Bénabou and 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 and Verrina, 2021; Messick and 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 and Weber, 2013).

[1]Note: our definition of self-serving bias is self-serving judgments of a fair division (Rodriguez-Lara and Moreno-Garrido, 2012; Cappelen et al., 2007). 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 (Deffains et al., 2016; Dorin et al., 2021; Miller and Ross, 1975; Bradley, 1978).

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, Sparkman et al. (2022) and Andre et al. (2021) find that most Americans are willing to support mitigation efforts, but they underestimate others' support for mitigation, undermining collective action. Findings from the experimental lab data on allocation decisions also suggest that these constructs have separate predictive power for behavior (Bašić and Verrina, 2021). Moreover, the extent to which different constructs predict behavior may depend on the strength of social image concerns and expectations of conformity (Bašić and Verrina, 2021; Thøgersen, 2008; Ajzen and Fishbein, 1970; Cialdini et al., 1991).

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 (Suhay et al., 2020; Valero, 2021; Deffains et al., 2016; Dorin et al., 2021; Cassar and Klein, 2019a; Di Tella et al., 2007). In contrast, those who are less successful or experience hardship are more likely to point to the role of luck or selfishness in success (Hvidberg et al., 2020; Hochleitner, 2022; Almås et al., 2022). 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. Cherry et al., 2002; Krawczyk, 2010; Cappelen et al., 2013; Durante et al., 2014; Lefgren et al., 2016; Cappelen et al., 2017; Bortolotti et al., 2017).

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. Valero (2021) and Lobeck (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., 2020). 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 Appendix 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 (Konow, 2000; Cappelen et al., 2007; Rodriguez-Lara and 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 receiver 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 serves primarily to illustrate 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. Figure 1 displays the timeline shared by the two experiments.

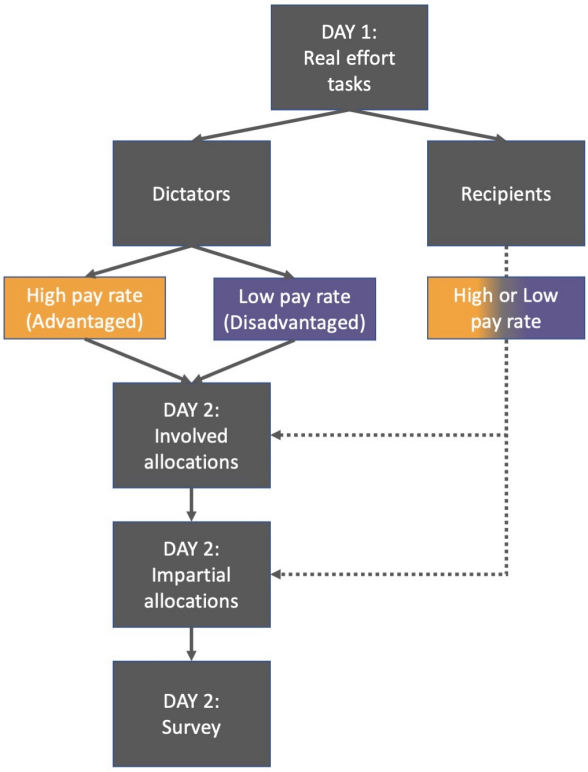
For Experiment 1, we recruited 200 dictators and 300 recipients from Prolific.co. The data was collected between the 13th and 19th of July, 2020. For Experiment 2, we recruited 400 dictators and 600 recipients from Prolific.co[2]. The data was collected between the 23rd and 30th of November, 2020. These sample sizes of 100 participants per treatment were preregistered (see Appendix C for preregistrations) and larger than those of similar studies (Konow, 2000; Rodriguez-Lara and Moreno-Garrido, 2012; Cappelen et al., 2007). 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. 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 randomly be assigned a high or low pay rate per correct answer, the amount of both pay

[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 in the Impartial trials the dictators split the amount generated by two recipients.

Figure 1: Timeline for Day 1 and 2 for Both Experiments.



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[3]. 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.

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"

[3]Participants could not move on in the experiment until they answered all the comprehension questions correctly.

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 "Privilege Status" or 'simply '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 [4].

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.[5]

[4] We control for purely mechanical carry-over effects in allocation by changing the orientation of the slider for half of the participants and include slider orientation in regressions looking at the impact of Status on allocation.

[5] 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.

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 and 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 [6]. 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 Appendices 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. Moreover, we elicited participants' demographics, including gender, country, political leaning, education, and income level.

[6] 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.

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. First, we asked participants to rate the moral appropriateness of dividing according to three fairness criteria that are commonly used in the literature (Konow, 2000; Cappelen et al., 2007; Rodriguez-Lara and 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). We use participants' appropriateness ratings as our main measure of personal norms.

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.

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.

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.

Our first question asked dictators for the number of trials they thought the recipient in the number of correct answers outperformed them. 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.[7]

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

[7] 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.

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). Figure 2 displays the share of the surplus that dictators allocated to the Advantaged member of the pair, split by Status treatment, and by Involved or Impartial allocation decisions.

Involved Allocations. Focusing on the Involved allocations, a rank-sum test of the average share each dictator gave to the Advantaged recipients across rounds confirms that the two groups allocate 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 give 10 percentage points more of the surplus to the Advantaged member ($p < 0.001$), an effect that is almost as large as the standard deviation of allocation decisions for this group.

The fact that Advantaged dictators allocate more to Advantaged members of the pair than Disadvantaged dictators do is consistent with dictators simply keeping most of the surplus. Therefore, we look at the impact of being Advantaged 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 keep more for themselves (Konow, 2000; Rodriguez-Lara and Moreno-Garrido, 2012; Deffains et al., 2016). 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. This relatively even division accords with previous work showing that dictators respect earned income in their allocations (Cappelen et al., 2010; Rodriguez-Lara and Moreno-Garrido, 2012; Cherry et al., 2002).

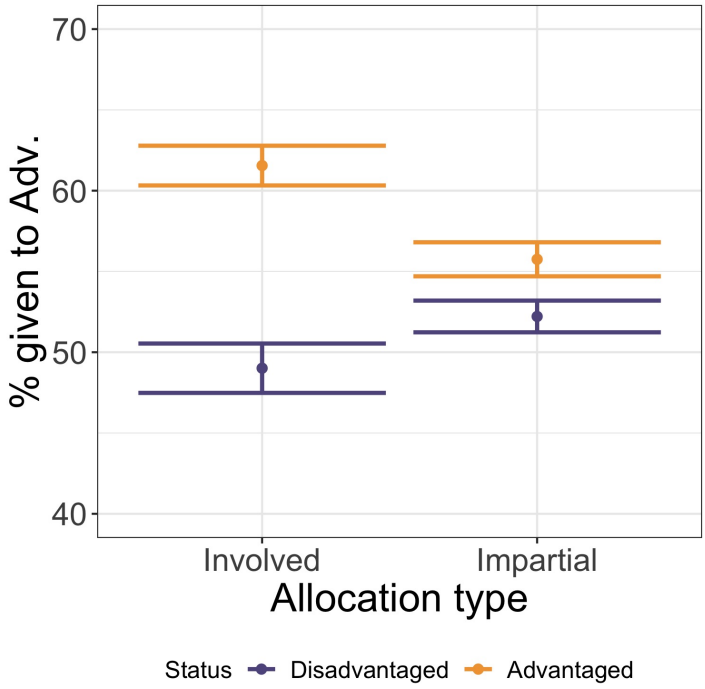
Impartial Allocations. In the Impartial allocation task, we removed the self-interest of the dictators. Any remaining favoritism towards the Advantaged dictators thus measures the per-

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.

Figure 2: Allocation by 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)	(2)	(3)
	% given to Adv.	% Kept	% 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. 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)). Clustered standard errors clustered at the individual level 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), task type (4 categories), slider orientation (2 categories).

sistence of a self-serving bias that self-interest cannot explain. The right part of Figure 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.*

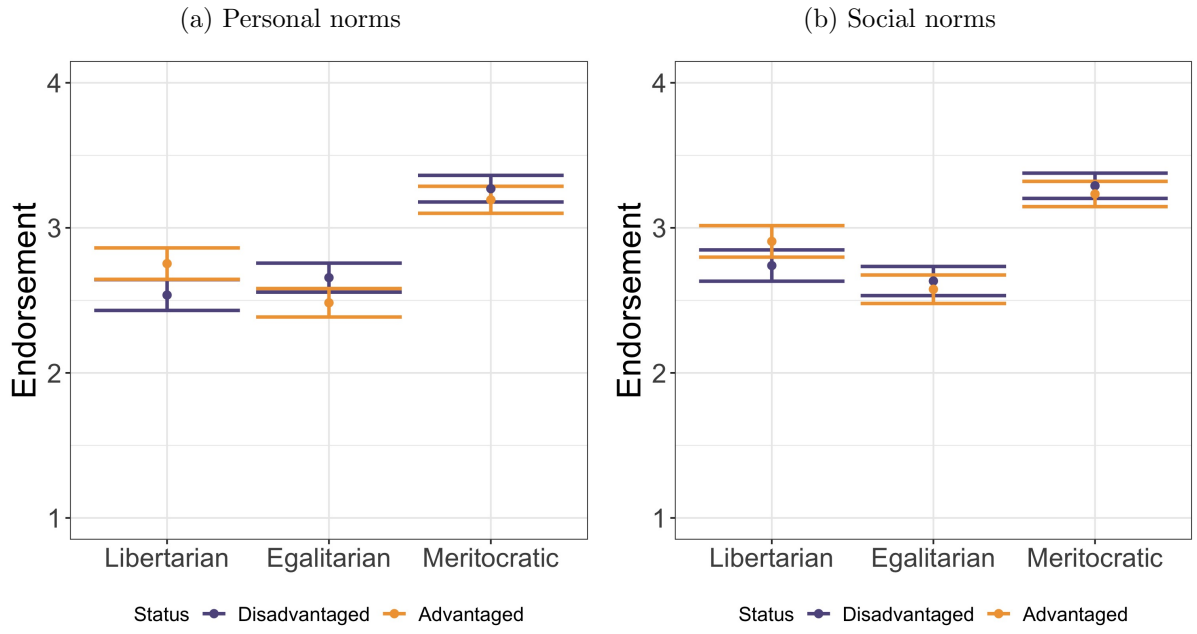
4.2 Status, Norms and Beliefs

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

Personal and Social Norms. We first look at both personal moral norms and anticipated social norms about the appropriateness of different fairness criteria. Figure 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.[8] 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 a difference in the distribution of answers for 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.

[8] Appendix A.1 investigates the effect of Status on our secondary norms elicitations. The results are qualitatively similar to the ones presented in this section but without statistical significance.

Figure 3: Personal and social norms by Treatment.



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

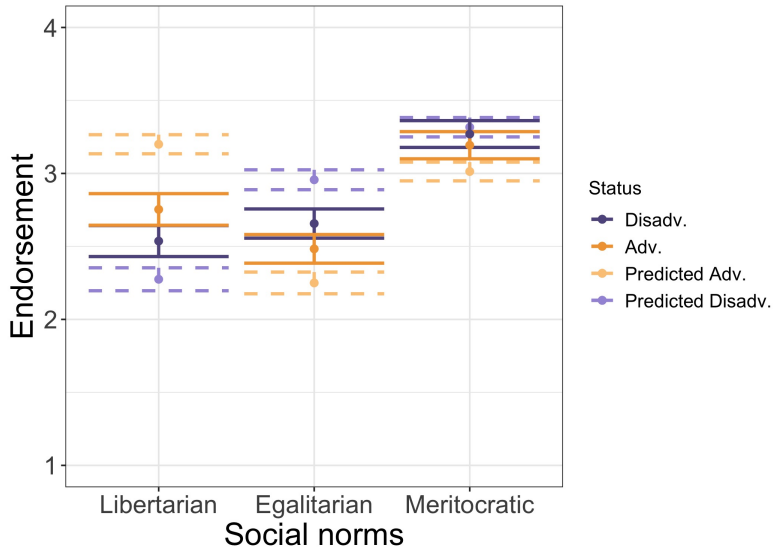
Data from Experiment 1 and Experiment 2. All models are ordered logits. Dependent variable: social or moral acceptability of a norm (1 very inappropriate, 2 somewhat inappropriate, 3 somewhat appropriate, 4 very appropriate). Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 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).

Can participants predict the effect of Status on norms? We investigate whether participants are 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 and 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 predict the Status gap in personal norms, we compare personal norms and predicted social norms in Figure 4. We find that participants, regardless of Status, correctly anticipate that Advantaged dictators endorse libertarian norms more highly (rank-sum tests $p < 0.001$) and Disadvantaged endorse 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 Figure 4, they overestimate the Status biases in social norms compared to the actual differences observed in personal norms and further predict that the Disadvantaged will be more likely to endorse meritocratic norms (rank-sum $p < 0.001$), suggesting that they anticipate others to have stronger status biases than themselves.

Despite the awareness of how status influenced self-serving biases in allocations, we do not

Figure 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.

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: ρ = -0.07, $p = 0.07$; meritocratic gap: ρ = -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.

Beliefs about relative performance. We now turn to beliefs about relative performance, elicited only in Experiment 2. Figure 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 $\sim$ 80% of Advantaged dictators indicate that the other participant did equally well or worse than them, compared to

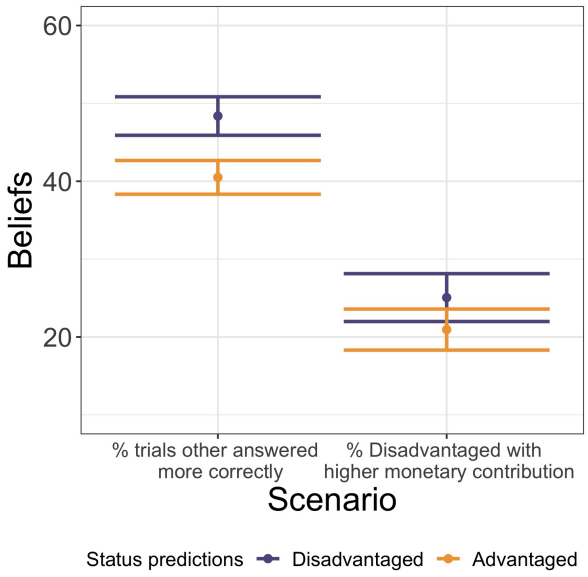
only 60% of Disadvantaged dictators. A rank-sum test confirms that these two groups' belief distributions are significantly different ($p < 0.001$). This result is also supported by an OLS regression displayed in Table 4, Column 1. It shows that Advantaged dictators believe that recipients answered more 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 (Valero, 2021; Deffains et al., 2016). Nevertheless, the difference in bias 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 by the beliefs about the probability that a Disadvantaged member would out-contribute an Advantaged member. Both groups of subjects overestimate this variable: the real chance is 6.8% while they believe it to be 23% (t-test, $p < 0.001$). In addition, we do not find support for the idea that Advantaged dictators underestimate their random advantage more to downplay the role of luck. A rank sum test shows no significant difference between the two groups' beliefs about monetary contributions ($p = 0.068$). 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 interpret larger monetary contributions as signifying a higher deservingness instead of a larger artificial advantage.[9]

Result 2. *Advantaged dictators find egalitarian sharing rules less and libertarian sharing rules personally more compelling than Disadvantaged dictators. We find similar but statistically non-significant results for overall social norms perceptions, and we show significant predictions of Status bias in social norms split by Dis(Advantage), regardless of one's Status. Advantaged dictators are also more likely to believe that they outperformed and out-contributed in the task.*

[9]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.

Figure 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.

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 and Kenny, 1981).

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 of beliefs and norms 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 of personal

Table 4: Effect of Status on beliefs

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

Data from Experiment 2. All models are linear regressions. 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. Clustered standard errors clustered at the individual level 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).

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

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). The parentheses under the coefficients for "Advantaged" report the standard errors clustered at the individual level; The parentheses below the check marks for report the $\chi^2(3)$ (for norms) and the $\chi^2(2)$ (for beliefs) statistics for the joint significance of these variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The F-statistics is the Kleibergen-Paap rk Wald F statistic.

and social norms. More precisely, we instrument our measures of personal norms with the alternative elicitation of the appropriateness of using different types of information to divide the common account, which measure the same 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.

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

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. The second stage shows that the coefficient for Advantaged drops to 1.76, is no longer significantly different from zero ($p = 0.090$), and it is significantly different from the from 3.04 coefficient in Column (1) ($t(372) = 2.83$), $p = 0.005$).[10] Column 3 performs the same analysis as in Column 2 but uses 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

[10]We compute this t-statistic using the method described in Clogg et al. (1995) to compare the coefficients of nested models. The similar tests below use the same approach.

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 norm, from Column (2).[11] 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.

Do norms and beliefs predict allocations? Next to the mediation, we look at the direct connection between norms and beliefs, and allocation. 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.[12] 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

[11] An alternative approach to compute the mediation effect is based on the product of the effect of Status on norms and of the effect of norms on allocation. This approach described by VanderWeele and Vansteelandt (2014) yields the same results as the one we have presented.

[12] Consistently with this approach, we don't display the individual coefficient in Table 5, the interested reader can find these coefficients in Appendix Table 6.

behavior ($\chi^2(2) = 2.27$, $p = 0.32$). Table 7 in the Appendix shows evidence that the effect of personal norms on behavior depends on the dictator's status.

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.[13]. An additional limitation is that the analysis assumes a linear relationship between norms, beliefs, and redistribution. To address these limitations, the regressions reported in Appendix 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.[14] The results of this alternative mediation analysis are similar to the one 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

This section discusses further aspects and interpretations of our main results. A number of design features of our study have implications for the generalizability and interpretation of our findings.

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.

[13] This weak first stage appears to be caused by collinearity in some of our instruments.

[14] Gillen et al. (2019) indicate that linearly including multiple measurements of the same variable can help to reduce measurement error.

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 it first.

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. 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 (Kallgren et al., 2000; Cialdini et al., 1991). 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 taken 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 doesn't 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 a 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 and 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 (Dorin et al., 2021; Cassar and Klein, 2019b). 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.

References

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

Almås, Ingvild, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden, "Global evidence on the selfish rich inequality hypothesis," *Proceedings of the National Academy of Sciences*, 2022, *119* (3), e2109690119.

Amasino, Dianna, Davide Pace, and Joël J. van der Weele, "Fair Shares and Selective Attention," *Tinbergen Discussion Paper*, 2021, *2021-066.*

Andre, Peter, Teodora Boneva, Felix Chopra, and Armin Falk, "Fighting climate change: The role of norms, preferences, and moral values," 2021.

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

_ , _ , Samuel Issacharoff, and Colin Camerer, "Biased Judgments of Fairness in Bargaining," *The American Economic Review*, 1995, *85* (5), 1337–1343.

Bašić, Zvonimir and Eugenio Verrina, "Personal Norms — and Not Only Social Norms — Shape Economic Behavior," SSRN Scholarly Paper ID 3720539, Social Science Research Network, Rochester, NY April 2021.

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Bicchieri, Cristina, Eugen Dimant, and Silvia Sonderegger, "It's not a lie if you believe it: On norms, lying, and self-serving belief distortion," Technical Report, CeDEx Discussion Paper Series 2019.

Blouin, Arthur and Sharun W. Mukand, "Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda," *Journal of Political Economy*, June 2019, *127* (3), 1008–1062.

Bortolotti, Stefania, Ivan Soraperra, Matthias Sutter, and Claudia Zoller, "Too Lucky to Be True - Fairness Views Under the Shadow of Cheating," SSRN Scholarly Paper ID 3014734, Social Science Research Network, Rochester, NY July 2017.

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

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, October 2020, *110* (10), 2997–3029.

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

_ , Erik Ø. Sørensen, and Bertil Tungodden, "Responsibility for what? Fairness and individual responsibility," *European Economic Review*, April 2010, *54* (3), 429–441.

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

_ , Karl O. Moene, Siv-Elisabeth Skjelbred, and Bertil Tungodden, "The Merit Primacy Effect," SSRN Scholarly Paper ID 2963504, Social Science Research Network, Rochester, NY April 2017.

Cassar, Lea and Arnd H. Klein, "A Matter of Perspective: How Failure Shapes Distributive Preferences," *Management Science*, 2019, *65* (11), 5050–5064.

_ and Arnd H Klein, "A matter of perspective: How failure shapes distributive preferences," *Management Science*, 2019, *65* (11), 5050–5064.

Charness, Gary, Ryan Oprea, and Sevgi Yuksel, "How do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment," *Journal of the European Economic Association*, June 2021, *19* (3), 1656–1691.

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

Cialdini, Robert B, Carl A Kallgren, and Raymond R Reno, "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, Elsevier, 1991, pp. 201–234.

Clogg, Clifford C., Eva Petkova, and Adamantios Haritou, "Statistical Methods for Comparing Regression Coefficients Between Models," *American Journal of Sociology*, March 1995, *100* (5), 1261–1293.

Cohn, Alain, Lasse J. Jessen, Marko Klasnja, and Paul Smeets, "Why Do the Rich Oppose Redistribution? An Experiment with America's Top 5%," SSRN Scholarly Paper ID 3395213, Social Science Research Network, Rochester, NY May 2019.

d'Adda, Giovanna, Michalis Drouvelis, and Daniele Nosenzo, "Norm elicitation in within-subject designs: Testing for order effects," *Journal of Behavioral and Experimental Economics*, 2016, *62*, 1–7.

Deffains, Bruno, Romain Espinosa, and Christian Thöni, "Political self-serving bias and redistribution," *Journal of Public Economics*, February 2016, *134*, 67–74.

Dengler-Roscher, Kathrin, Natalia Montinari, Marian Panganiban, Matteo Ploner, and Benedikt Werner, "On the malleability of fairness ideals: Spillover effects in partial and impartial allocation tasks," *Journal of Economic Psychology*, 2018, *65*, 60–74.

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

Durante, Ruben, Louis Putterman, and Joël van der Weele, "Preferences for Redistribution and Perception of Fairness: An Experimental Study," *Journal of the European Economic Association*, August 2014, *12* (4), 1059–1086.

Espinosa, Romain, Bruno Deffains, and Christian Thöni, "Debiasing preferences over redistribution: an experiment," *Social Choice and Welfare*, December 2020, *55* (4), 823–843.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," *Journal of Political Economy*, June 2019.

Hochleitner, Anna, "Fairness in times of crisis: Negative shocks, relative income and preferences for redistribution," Technical Report, CeDEx Discussion Paper Series 2022.

Hvidberg, Kristoffer B., Claus Kreiner, and Stefanie Stantcheva, "Social Position and Fairness Views," Technical Report, National Bureau of Economic Research 2020.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto, "Experimental designs for identifying causal mechanisms," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2013, *176* (1), 5–51.

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

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

Konow, James, "Fair shares: Accountability and cognitive dissonance in allocation decisions," *American Economic Review*, 2000, *90* (4), 1072.

Koo, Hyunjin J., Paul K. Piff, and Azim F. Shariff, "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*, 2022, *0* (0), 19485506221098921.

Krawczyk, Michał, "A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution," *Journal of Public Economics*, February 2010, *94* (1), 131–141.

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

Lefgren, Lars J., David P. Sims, and Olga B. Stoddard, "Effort, luck, and voting for redistribution," *Journal of Public Economics*, November 2016, *143*, 89–97.

Lobeck, Max, "Motivating Beliefs in a Just World," 2021.

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

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

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

Rodriguez-Lara, Ismael and Luis Moreno-Garrido, "Self-interest and fairness: self-serving choices of justice principles," *Experimental Economics*, 2012, *15* (1), 158–175.

Rustichini, Aldo and Marie Claire Villeval, "Moral hypocrisy, power and social preferences," *Journal of Economic Behavior & Organization*, 2014, *107*, 10–24.

Saccardo, Silvia and Marta Serra-Garcia, "Enabling or Limiting Cognitive Flexibility? Evidence of Demand for Moral Commitment," *American Economic Review*, 2023, *113* (2), 396–429.

Sandel, Michael J., *The Tyranny of Merit: What's become of the common good?*, Allen Lane London, 2020.

Schlag, Karl H., James Tremewan, and Joël J. Van der Weele, "A penny for your thoughts: A survey of methods for eliciting beliefs," *Experimental Economics*, 2015, *18* (3), 457–490.

Schwardmann, Peter, Egon Tripodi, and Joël J. Van der Weele, "Self-persuasion: Evidence from field experiments at international debating competitions," *American Economic Review*, 2022, *112* (4), 1118–1146.

Sparkman, Gregg, Nathan Geiger, and Elke U Weber, "Americans experience a false social reality by underestimating popular climate policy support by nearly half," *Nature communications*, 2022, *13* (1), 1–9.

Suhay, Elizabeth, Marko Klasnja, and Gonzalo Rivero, "Ideology of Affluence: Rich Americans' Explanations for Inequality and Attitudes toward Redistribution," 2020.

Tella, Rafael Di, Sebastian Galiant, and Ernesto Schargrodsky, "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters," *The Quarterly Journal of Economics*, February 2007, *122* (1), 209–241.

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

Ubeda, Paloma, "The consistency of fairness rules: An experimental study," *Journal of Economic Psychology*, April 2014, *41*, 88–100.

Valero, Vanessa, "Redistribution and beliefs about the source of income inequality," *Experimental Economics*, September 2021.

VanderWeele, Tyler and Stijn Vansteelandt, "Mediation analysis with multiple mediators," *Epidemiologic methods*, 2014, *2* (1), 95–115. Publisher: De Gruyter.

Willemsen, Martijn C. and Eric J. Johnson, "Observing Cognition with MouselabWEB," in "A handbook of process tracing methods" 2019, pp. 76–95.

Appendices

A Additional results

A.1 The effect of Status on alternative measures of personal norms

In this section we study the effect of Status on personal norms using the alternative/secondary measures of these norms. While in Section 4.3 we measured the personal norms with questions about the fairness of certain allocations, we now focus on how dictators describe their choices in an open-ended question, dictators' beliefs about the morality of using different types of information as input for redistribution, and attitudes towards redistribution in the real world. The aim of the analysis is to test the robustness of the effects we found in Section 4.3 and better characterize the relationship between Status and personal norms.

First, we look at the open-ended answers to the question: "How did you decide how to split the common account?". An RA classified the fairness views indicated in the answers. The RA was oblivious to the research question. Table 1 counts the mentions of a specific fairness criterion by dictators split by Status, regardless of whether their behavior was in line with this criterion. We do not count dictators who mention other criteria or who mention multiple criteria.

We find that slightly more than half of the 600 dictators claim to be meritocratic, i.e., to distribute on the basis of the correct answers, and that this fraction does not differ based on Status ($p = 0.07$, two-sided Fisher's exact test). Smaller groups say they used egalitarian splits or libertarian divisions (according to the monetary contribution). Among these last groups, we see a status divide in the direction one expects: Advantaged dictators have about half the number of egalitarians ($p = 0.04$, two-sided Fisher's exact test) and twice the number of libertarians ($p = 0.06$, two-sided Fisher's exact test). Taken together, these patterns line up with an overall interpretation that economic position affects norms of fairness.

Next, we look at beliefs about the fairness of redistributing money using different types of information. In the first question, we asked dictators how morally appropriate they consider splitting the money exclusively using information about the monetary contribution. Using this information would result in a libertarian split, and indeed there is a significant and positive correlation between the answers to this question and the beliefs about the moral appropriateness of a libertarian allocation ($\rho = 0.41$, $p < 0.001$). Column 1 of Table 2 uses an ordered logistic regression to show that there is no difference in Status in the likelihood of considering using information about the monetary contribution morally appropriate ($p = 0.17$). In a second

Table 1: Fairness criteria in open-ended questionnaire item.

	Meritocratic	Egalitarian	Libertarian
Disadvantaged	170	25	15
Advantaged	147	12	28
Total	317	37	43

In this table, we count the number of subjects per treatment that, in an open-ended question, mention that they redistributed the money according to one of the meritocratic, egalitarian, or libertarian criteria. We do not count participants that mention more than one criterion in their answers.

question, we asked dictators how morally appropriate they consider redistributing exclusively using information about the number of correct answers in the task. Using this information produces a meritocratic split and, once again, there is a significant and positive correlation between the answers to this question and the moral appropriateness rating of a meritocratic allocation ($\rho = 0.38$, $p < 0.001$), but no difference in Status (Column 2 of Table 2 ($p = 0.31$). The lack of significant results may be because the question about the use of certain information is more abstract or indirect than simply asking for ratings of fairness criteria. However, the direction of the effects is consistent with the ones found in Section 4.3.

Finally, we study attitudes towards redistribution outside the experiment. Here we ask participants whether they think that the government should take measures to reduce differences in income levels. Column 3 of Table 2 shows no difference by Status on redistribution in society ($p = 0.28$).

Table 2: The effect of Status on personal norms - secondary elicitations

	(1) usingContribution	(2) usingAnswers	(3) redistribution
Advantaged	0.26	-0.21	0.18
	(0.19)	(0.20)	(0.17)
Observations	400	400	600

Data from Experiment 2. All models are ordered logits. Dependent variables: (1) morality of using exclusively information about participants' monetary contributions to determine the allocation; (2) morality of using exclusively information about participants' number of correct answers to determine the allocation; (3) Agreement with the statement "The government should take measures to reduce differences in income levels". 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).

A.2 The importance of correcting for measurement error

In this section, we quantify the importance of correcting for measurement error when studying the psychological antecedents of self-serving biases. To do so, we compare the results of our IV estimation with an OLS estimation that linearly controls only for our main norms elicitations. This second and simpler approach is the one followed by Dorin et al. (2021) and by most mediation analyses in the literature.

Table 3 in the Appendix presents the results of this analysis. Column 1 displays our usual benchmark for self-serving biases indicating that the allocations made by Advantaged and Disadvantaged dictators differ by 3.04 percentage points. Column 2 adds linear controls for personal norms without accounting for measurement error. Doing so reduces the coefficient for the Advantaged dummy to 2.52, indicating that the variation in personal norms explains 17% of self-serving biases. Using a similar approach, Dorin et al. (2021) find that personal norms explain 16% of self-serving biases, an estimate very close to ours. However, comparing these estimates with Column 2 of Table 5 in the main text shows that not accounting for measurement error leads to underestimating the explanatory power of personal norms by a factor of 2.5.

Column 3 repeats the same analysis as in Column 2 for social norms. Here we find that linearly controlling for social norms using a single set of elicitations explains 11% of self-serving biases. This mediation effect is close to the one we found in Column 3 of Table 5 of the main text using the IV approach. The similarity of the results suggests that measurement error is less of a concern for social norms, possibly because this kind of norm can be incentivized (Schlag et al., 2015).

We conclude that addressing measurement error is essential to correctly quantify the role of norms as psychological antecedents of self-serving biases, in particular for variables that cannot be incentivized.[15]

A.3 Estimating the joint mediating role of norms and beliefs.

The empirical mediation strategy presented in Section 4.3 (Table 5 in the main text) has two limitations. First, it assumes a linear relationship between norms, beliefs, and redistribution. Second, the instrumental variables approach does not allow us to estimate the joint mediating effect of personal norms, social norms, and beliefs.

To address these limitations, the regressions reported in Table 4 control for beliefs entering

[15]Table 5 in this Appendix shows that including personal norms and social norms as dummy variables rather than linearly as in Table 3 produces similar results.

Table 3: Allocation to the Advantaged controlling for norms and beliefs without correcting for measurement error

	(1)	(2)	(3)
	% given to Adv.	% given to Adv.	% given to Adv.
Advantaged	3.04***	2.52**	2.71**
	(0.86)	(0.86)	(0.86)
Personal Norms		✓	
Social Norms			✓
Observations	400	400	400

Data: Impartial trials from Experiment 2. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. Personal Norms: perLib, perEga, perMer. Social Norms: socLib, socEga, socMer. 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 clustered at the individual level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

them as 4th-degree polynomials and for personal and social norms entering them as dummy variables for each possible norm rating. Moreover, to limit biased estimates due to measurement error, every set of elicitations available for the norms is included in the regression.[16] In total, we use 6 categorical variables (24 dummies) for the personal norms and 9 categorical variables (36 dummies) for the social norms.

Table 4 shows that this alternative approach leads to similar results as the one discussed in Section 4.3. Column 1 is the same as in Table 5 in the main text, and it is reported here to facilitate the comparisons. Column 2 shows that the personal norms explain 38% of self-serving biases - slightly less than under our IV approach. Column 3 shows that social norms do not contribute to explaining self-serving bias. Column 4 indicates that the explanatory power of beliefs does not increase if we allow for nonlinear effects. Finally, Columns 5 and 6 show that our ability to explain self-serving biases does not change much if we control for personal norms, social norms, and beliefs jointly. This analysis thus confirms that shifts in personal norms are the important determinant of self-serving biases and that social norms and beliefs play a secondary role if any.

[16] Gillen et al. (2019) indicate that linearly including multiple measurements of the same variable can help to reduce measurement error

Table 4: Allocation to Advantaged recipients controlling for norms and beliefs

	(1) % given to Adv.	(2) % given to Adv.	(3) % given to Adv.	(4) % given to Adv.	(5) % given to Adv.	(6) % given to Adv.
Advantaged	3.04^{***}	1.89^{*}	3.09^{***}	2.99^{**}	2.07^{*}	1.78
	(0.86)	(0.81)	(0.86)	(0.94)	(0.85)	(0.98)
Personal Norms		✓			✓	✓
Social Norms			✓		✓	✓
Beliefs				✓		✓
Observations	400	400	400	400	400	400

Data: Impartial trials from Experiment 2. Dependent variable: percentage of the surplus allocated to the Advantaged member of the pair. Personal Norms: a dummy variable for each possible value of each variable of our three sets personal norms elicitation (morality of splitting according to a fairness criterion, morality of using different types of information, mentioning a fairness criterion in the open-ended question). Social Norms: dummy variable for each possible value of each variable of our three social norms elicitations (norms of the universe of dictators, of the Advantaged dictators, and of the Disadvantaged dictators). Beliefs: # RecOutperf and % DisOutcont; both beliefs enter the regressions as polynomials of degree 4. # DisOutperfom 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". 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 clustered at the individual level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.4 Robustness check: alternative specification to show the importance of controlling for measurement error

Table 5: Impartial allocations to Advantaged recipients controlling for norms and beliefs without controlling for measurement error.

	(1)	(2)	(3)
	% given to Adv.	% given to Adv.	% given to Adv.
Advantaged	3.04***	2.52**	2.79**
	(0.86)	(0.85)	(0.85)
Personal Norms		✓	
Social Norms			✓
Observations	400	400	400

Data: Impartial trials from Experiment 2. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. Personal Norms: a dummy variable for each possible value of perLib, perEga, perMer. Social Norms: dummy variable for each possible value of each variable of socLib, socEga, socMer. 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 clustered at the individual level in parentheses.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.5 The effects of norms and beliefs on allocation

Table 6: Impartial allocations to Advantaged recipients controlling for norms and beliefs

	(1)	(2)	(3)	(4)
	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.
Advantaged	3.04***	1.75	2.71**	2.82**
	(0.84)	(1.06)	(0.85)	(0.94)
perLib		6.80***		
		(1.61)		
perEga		-0.026		
		(1.53)		
perMer		-0.75		
		(2.17)		
socLib			1.57*	
			(0.66)	
socEga			0.070	
			(0.94)	
socMer			-1.85+	
			(1.06)	
# DisOutperf				-0.14
				(0.14)
% DisOutcont				0.020
				(0.019)
F-statistic		10.8	20.9	
Observations	7939	7939	7939	7939

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 a) the Advantaged dictators and b) the Disadvantaged dictators. In Column (4), "# DisOutperfom" 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". All the other variables mentioned in the table 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 clustered at the individual level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The F-statistics is the Kleibergen-Paap rk Wald F statistic.

Table 7: Allocation to the Advantaged controlling for norms and beliefs, and their interaction with status

	(1)	(2)	(3)
	% given to Adv.	% given to Adv.	% given to Adv.
Advantaged	-2.03	-9.68	2.59
	(4.50)	(6.10)	(3.02)
perLib	0.22		
	(0.58)		
perEga	-0.32		
	(0.52)		
perMer	-1.98**		
	(0.61)		
Adv*perLib	2.47**		
	(0.77)		
Adv*perMer	0.020		
	(0.87)		
Adv*perEga	-0.66		
	(0.76)		
socLib		0.32	
		(0.63)	
socEga		0.036	
		(0.65)	
socMer		-3.03**	
		(0.95)	
Adv*socLib		1.94*	
		(0.87)	
Adv*socMer		1.96	
		(1.33)	
Adv*socEga		0.20	
		(0.94)	
# DisOutperform			-0.15
			(0.20)
% DisOutcont			0.018
			(0.024)
Adv*# DisOutperform			0.013
			(0.30)
Adv*% DisOutcont			0.0047
			(0.041)
Observations	600	400	400

Data: Impartial trials from Experiment 2. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. "# DisOutperfom" 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". All the other variables mentioned in the table are defined in Table 1 of the main text. 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 clustered at the individual level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.6 Robustness check: excluding rounds in which participants are not fully attentive

As we detailed in Section 3 and the companion paper Amasino et al. (2021), our experimental design aimed to identify the relationship between participants' attention and redistributive decisions. In particular, Experiment 2 implemented some variations designed to shift attention to information about the origin of the surplus. An obvious concern is that this attention manipulation drives our results.

Here, we replicate our mediation analysis but exclude the rounds in which participants are not fully attentive. In the experiment, dictators might decide not to access all the available information in some or all the rounds, and which information they decide to avoid might be correlated with the treatment they are in. To factor out selective attention, we created a database that includes only the rounds in which the dictators accessed all the available information. In addition, we collapsed the database at the dictator level to make sure that every dictator gets the same weight in the analysis. This database contains 384 dictators from Experiment 2, 96% of the original sample.

Table 8 and Table 9 use this restricted database to replicate, respectively, the IV specification (from Section 4.3) and the dummy variables specifications (from Appendix A.3) for the mediation analysis. The tables show that personal norms explain between 28% and 31% of the effect of status on allocation. The tables also show that a) the explanatory power of social norms remains small and insignificantly different from zero, and b) the explanatory power of beliefs is between 15% and 16%, somewhat higher than in the main analysis. In addition, Table 10 reinforces the importance of accounting for measurement error. It uses the restricted database to conduct a mediation analysis in which the norms enter as linear controls without being instrumented. The results confirm that this simpler approach underestimates the mediating effect of personal norms by a factor of 2.5.

Table 8: Impartial allocations to Advantaged recipients controlling for norms and beliefs - only trials where participants are fully attentive

	(1)	(2)	(3)	(4)
	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.
Advantaged	3.33***	2.40*	3.10**	2.78**
	(0.97)	(1.14)	(1.05)	(1.06)
perLib		6.04***		
		(1.71)		
perEga		0.23		
		(1.87)		
perMer		-0.97		
		(2.20)		
socLib			1.01	
			(0.81)	
socEga			0.14	
			(1.14)	
socMer			-1.28	
			(1.51)	
# DisOutperfom				-0.24
				(0.15)
% DisOutcont				-0.0050
				(0.027)
F-statistic		10.8	18.6	
Observations	384	384	384	384

Data: Impartial trials from Experiment 2, trials where participants do not access all the information are excluded, and the database is collapsed at the dictator level. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. Columns (1) and (4) are linear regressions, while 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' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. List of controls common to all regressions: percentage of the joint number of correct answers due to the Advantage 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). Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The F-statistics is the Kleibergen-Paap rk Wald F statistic.

Table 9: Allocations to Advantaged recipients controlling for norms and beliefs - only trials where participants are fully attentive

	(1) % given to Adv.	(2) % given to Adv.	(3) % given to Adv.	(4) % given to Adv.	(5) % given to Adv.	(6) % given to Adv.
Advantaged	3.33**	2.31*	3.51***	2.82*	2.61*	1.90
	(1.01)	(1.00)	(1.05)	(1.09)	(1.07)	(1.17)
Personal Norms		✓			✓	✓
Social Norms			✓		✓	✓
Beliefs				✓		✓
Observations	384	384	384	384	384	384

Data: Impartial trials from Experiment 2, trials where participants do not access all the information are excluded, and the database is collapsed at the dictator level. Dependent variable: percentage of the surplus allocated to the Advantaged member of the pair. Beliefs: % DisOutcont and # DisOutperfom; both beliefs enter the regressions as polynomials of degree 4. # DisOutperfom 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". Personal Norms: a dummy variable for each possible value of each variable of our three sets personal norms elicitation (morality of splitting according to a fairness criterion, morality of using different types of information, mentioning a fairness criterion in the open-ended question). Social Norms: dummy variable for each possible value of each variable of our three social norms elicitations (norms of the universe of dictators, of the Advantaged dictators, and of the Disadvantaged dictators) . 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). Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 10: Allocations to the Advantaged recipients controlling for norms and beliefs without correcting for measurement error - only trials in which participants are fully attentive

	(1) % given to Adv.	(2) % given to Adv.	(3) % given to Adv.
Advantaged	3.33**	2.97**	3.07**
	(1.01)	(1.04)	(1.03)
Personal Norms		✓	
Social Norms			✓
Observations	384	384	384

Data: Impartial trials from Experiment 2, trials where participants do not access all the information are excluded, and the database is collapsed at the dictator level. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. Personal Norms: perLib, perEga, perMer. Social Norms: socLib, socEga, socMer. 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). Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.7 Robustness check: including the data from Experiment 1 in the mediation analysis

We further test the robustness of our mediation analysis to attentional effects by running additional regressions that include data from Experiment 1, which did not include any attention manipulations. We excluded this experiment from our main analysis in Table 5, because in this experiment we did not elicit the performance beliefs or the secondary elicitations for personal norms. For this reason, in the analysis we discuss next, we use the dummies approach to control for measurement error of personal norms, as we did in Table 4.

Table 11 combines the data from Experiment 1 and 2, while Table 12 uses only the data from Experiment 1. These robustness checks show that personal norms explain 32% and 23% of self-serving biases, in Column 4 of Table 11 and Table 12 respectively, compared to 38% in Table 4 and 42% in Table 5. This discrepancy likely stems from additional measurement error due to the lack the secondary elicitations. When it comes to social norms, where we can still use the IV approach, we find that they explain explain 12% and 8% of self-serving biases, in Column 3 of Table 11 and Table 12 respectively. These numbers are similar to our finding of 11% in Table 5.

Taking Appendices A.6 and A.7 together, these additional checks provide evidence that attention does not play a large role in influencing norms and that our attention manipulation does not confound our results.

Table 11: Allocations to the Advantaged recipients controlling for personal and social norms - all data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.
Advantaged	3.41***	1.21	2.99***	2.32***	3.32***	2.33	2.48***
	(0.71)	(1.89)	(0.69)	(0.65)	(0.70)	(1.45)	(0.69)
IV for personal norms		✓				✓	
Personal norms dummies				✓			✓
IV for social norms			✓			✓	
Social norms dummies					✓		✓
F-statistic		0.003	33.3			0.196	
Observations	600	600	600	600	600	600	600

All data from the Impartial trials. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. In Column 2, the model is a 2SLS; the instrumented variables are perLib, perEga, and perMer; the instruments are dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms. In Column 3, the model is a 2SLS; the instrumented variables are socLib, socMer, and socEga; the instruments are the dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. In Column 4, the model is a linear regression; the model includes dummies for all the possible values of perLib, perEga, perMer, and dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms. In Column 5, the model is linear regression; the model includes dummies for all the possible values of socLib, socEga, socMer, and dummies for all the possible values of dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. In Column 6, the model is a 2SLS; the instrumented variables are perLib, perEga, perMer, socLib, socMer, and socEga; the instruments are dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms, and the dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. In Column 7, the model is a linear regression; the model includes dummies for all the possible values of perLib, perEga, perMer socLib, socEga, socMer, dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms, and dummy variables for dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. List of controls common to all models: 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 clustered at the individual level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The F-statistics is the Kleibergen-Paap rk Wald F statistic.

Table 12: Allocations to the Advantaged recipients controlling for personal and social norms - Experiment 1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.	% given to Adv.
Advantaged	3.90**	1.94	3.62**	2.93*	3.88**	2.92	2.50
	(1.33)	(2.64)	(1.29)	(1.19)	(1.42)	(2.17)	(1.48)
IV for personal norms		✓				✓	
Personal norms dummies				✓			✓
IV for social norms			✓			✓	
Social norms dummies					✓		✓
F-statistic		0.6	6.5			0.2	
Observations	200	200	200	200	200	200	200

Data from the Impartial trials of Experiment 1. Dependent variables: percentage of the surplus allocated to the Advantaged member of the pair. In Column (2), the model is a 2SLS; the instrumented variables are perLib, perEga, and perlMer; the instruments are dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms. In Column (3), the model is a 2SLS; the instrumented variables are socLib, socMer, and socEga; the instruments are the dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. In Column (4), the model is a linear regression; the model includes dummies for all the possible values of perLib, perEga, perMer, and dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms. In Column (5), the model is linear regression; the model includes dummies for all the possible values of socLib, socEga, socMer, and dummies for all the possible values of dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. In Column (6), the model is a 2SLS; the instrumented variables are perLib, perEga, perMer, socLib, socMer, and socEga; the instruments are dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms, and the dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. In Column (7), the model is a linear regression; the model includes dummies for all the possible values of perLib, perEga, perMer socLib, socEga, socMer, dummy variables indicating whether the dictator mentioned the different fairness criteria in the open-ended answer about personal norms, and dummy variables for dictators' beliefs about the social norms of a) the Advantaged dictators and b) the Disadvantaged dictators. List of controls common to all models: 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 clustered at the individual level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The F-statistics is the Kleibergen-Paap rk Wald F statistic.

B Illustrative model

We incorporate fairness ideals into redistributive choices, following the approach in Konow (2000) and Cappelen et al. (2007). Our model augments these previous approaches by looking at the (self-serving) determinants fairness ideals.

We consider a decision maker who has to allocate a monetary surplus of size X, by dividing it between her own share y and the share of a recipient $X - y$. She has utility function given by

$$u = y - \beta \frac{(y - \Phi(a))^2}{2X},$$

where $\Phi(a) \in [0, X]$ is the perceived fair share of the decision maker according to the decision maker, which may depend on her position in society a, i.e. advantaged or disadvantaged. Thus, the decision maker cares about her monetary payoffs y and suffers a psychological cost or guilt (scaled by β) when she deviates from her subjective fairness assessment.

Assuming an interior solution exists, u reaches a maximum at $y^* = \frac{X}{\beta} + \Phi(a)$, which is increasing in the size of the pie and the subjective fair share for the dictator, and decreasing in guilt sensitvity β.

Our paper asks where such fairness concerns come from, and how they depend on societal advantage of the decision maker. To this end, we operationalize $\Phi(a)$. Following the literature (e.g. Cappelen et al. (2007)), we assume that there are three possible fairness criteria: egalitarian, meritocratic and libertarian, denoted by Φ_E, Φ_M and Φ_L respectively, which each prescribe a particular split (see main text). The decision-maker's fairness concern is a weighted average of these three concerns

$$\Phi = \tilde{w}_E(a) * \Phi_E + \tilde{w}_M(a) * \Phi_M + \tilde{w}_L(a) * \Phi_L,$$

where $\tilde{w}_k(a)$ is the (normalized) weight on the different fairness criteria for $k \in \{E, M, L\}$, which may again depend on societal advantage a. To make sure that weights sum to 1 (and Φ cannot exceed X), they are normalized, i.e. $\tilde{w}_k(a) = \frac{w_k}{\sum_k w_k}$. The weight w_k in turn depends on several factors, namely attitudes or personal norms, perceived social norms and contextual beliefs

$$w_k = f(PN_k(a), SN_k(a), \{CB_k(a)\}),$$

where these components represent

- PN_k: Personal norms as measured by the appropriateness score.
- SN_k: Perceived social norms as measured by the appropriateness score.
- CB_k: Contextual beliefs that affect the applicability or relevance of the criterion. For instance, if one believes that there is a level playing field (our first elicitation), this may raise the weight on the meritocratic criterion. Furthermore, more optimistic beliefs about own relative performance may affect how much weight to place on meritocratic aspects of fairness.

For the sake of our analysis, we will assume that $f(\cdot)$ is separable and additive in all its components. Moreover, we abstract from how the precise measures are scaled, so we will not be able to make precise quantitative predictions. These limitations can be the subject of future research.

To see the role of self-serving bias, consider the impact of societal advantage a. If a increases, self-serving bias will increase the weight of a criterion that gives the highest payoff, i.e.. the libertarian criterion, and reduce the weight on other criteria. The model elucidates the channels through which such changes may take place: it may spur changes in personal norms of appropriateness, perceptions of social norms, or it may change other contextual beliefs like those about initial conditions. Through this change in weight, we will see an increase in Φ, the fair share of the dictator. This in turn lowers the psychological cost of keeping a larger share for the self, and translates into more selfish decisions.

C Preregistrations

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Tracking fairness (#44417)

Created: 07/12/2020 01:54 AM (PT)
Shared: 11/03/2020 12:17 PM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
We study the origin of self-serving biases in monetary allocation problems. If people are randomly placed in a (dis)advantaged position, how does this affect their attention to meritocratic information, the ethical criteria for making decisions, and the subsequent allocation choices? Detailed hypotheses are specified in point 5).

3) Describe the key dependent variable(s) specifying how they will be measured.
In Part 1 of the experiment, subjects first produce a surplus together with a matched partner on several tasks. We create variation in contribution to the surplus by randomly giving one of the partners a higher piece rate than the other. In Part 2 of the experiment, some subjects are given information on the performance on the tasks as well as the total contribution, and make allocation decisions in the role of dictator. We use Mouselab to track the way subjects explore information about task performance.

Per every decision of the dictator we record:
- the split in the total surplus between dictator and recipient.
- dwelling time (mousetracked) on each of the following information 1) the dictator & recipient contribution to the pie in monetary terms, 2) the number of answers in the task the dictator & recipient got correct.

4) How many and which conditions will participants be assigned to?
Subjects are assigned to be “receivers” and “dictators”. Both groups take part in a series of performance tasks to determine the surplus. We are mostly interested in the dictators.

All dictators are assigned to one of two treatments:
Advantaged: receives a high piece rate per correct answer in the task.
Disadvantaged: receives a low piece rate per correct answer in the task.

Each dictator participates (in this order) in an
Involved condition: 20 allocations between themselves and another randomly matched participant
Benevolent condition: 20 allocations between two other participants.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Hypothesis 1 (Behavior): In the involved condition, advantaged dictators give less money to the receivers than disadvantaged dictators.

We test this hypothesis with a non-parametric rank sum test. We will perform regressions to control for subject characteristics with standard errors clustered for each participant.

Hypothesis 2 (Attention): In the involved condition, advantaged dictators spend relatively less time on correct answer information and more time on monetary contribution information than disadvantaged dictators.
Across dictator groups, we investigate total time looking at information as well the proportion of time spent looking at correct answers, using a non-parametric rank sum test. We will also perform regressions with standard errors clustered for each participant.

Hypothesis 3 (Persistence): The effects documented in 1) and 2) persist in the benevolent condition.
The tests are the same as for Hypothesis 1 and 2, but now in the benevolent condition. We will also compare the effects in both conditions using a difference in difference approach.

Hypothesis 4 (Role of attention): Attention patterns drive giving decisions.
For correlational evidence, we use regressions to investigate how sensitive the treatment effect (Hypothesis 1) is to controlling for total and relative looking time. For a causal inference, we use an instrumental variable analysis to exploit variation generated by the (randomly varied) orientation of patterns on the

screen.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Following standard Mouselab protocols, we will exclude information that was revealed for less than 200 ms.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will recruit 200 dictators from the online platform Prolific. These are divided 50-50 between the advantaged and disadvantaged condition. We recruit the corresponding number of recipients.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will conduct a number of secondary analyses:
- We will compare by treatment the fairness criteria people list in the questionnaire as being most socially appropriate.
- We compare by treatment the fairness "types" based on Cappelen et al. (2007), and correlate these types with attentional patterns.
- Correlate attention, behavior and political preferences elicited in the final questionnaire.
In addition, we will explore additional measures of attention, and their explanatory power for giving decisions. We will conduct robustness analysis on the revelation threshold in point 6).

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Tracking fairness - attention manipulation (#52512)

Created: 11/18/2020 09:42 AM (PT)
Shared: 02/16/2021 06:11 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
We study the origin of self-serving biases in monetary allocation problems. If people are randomly placed in a (dis)advantaged position, how does this affect their attention to meritocratic information, the ethical criteria for making decisions, and the subsequent allocation choices? In a previous version of the experiment, we showed that advantaged dictators pay less attention to information that reveals pure merit (actual task performance). In this experiment we ask how randomly induced variations in attention affect decision making.

3) Describe the key dependent variable(s) specifying how they will be measured.
In Part 1 of the experiment, subjects first produce a surplus together with a matched partner on several tasks. We create variation in contribution to the surplus by randomly giving one of the partners a higher piece rate than the other. In Part 2 of the experiment, some subjects are given information on the performance on the tasks as well as the total contribution, and make allocation decisions in the role of dictator. We manipulate how long different kinds of information are available to people.

Per every decision of the dictator we record:
- the split in the total surplus between dictator and recipient.
- dwelling time (mousetracked) on each of the following information 1) the dictator & recipient contribution to the pie in monetary terms, 2) the number of answers in the task the dictator & recipient got correct.

4) How many and which conditions will participants be assigned to?
Subjects are assigned to be “receivers” and “dictators”. Both groups take part in a series of performance tasks to determine the surplus. We are mostly interested in the dictators.

All dictators are assigned to one of two treatments:
Advantaged: receives a high piece rate per correct answer in the task.
Disadvantaged: receives a low piece rate per correct answer in the task.

We cross-randomize these treatments with another dimension:
Merit focus: in a majority of trials, the information about task performance (merit) is available longer.
Output focus: in a majority of trials, information about total contribution to surplus is available longer.

Each dictator participates (in this order) in an
Involved condition: 20 allocations between themselves and another randomly matched participant
Benevolent condition: 20 allocations between two other participants

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We test two main hypotheses for both the involved and the benevolent dictators:
1) Dictators in the “Merit Focus” treatment will give more to disadvantaged recipients.

We will test this in a regression with data for all trials and a dummy for all trials with Merit Focus, as well as controls for subject and trial characteristics.

2) Compared to a situation with freely chosen attention, making dictators look longer at “inconvenient” information (i.e. ”Merit focus” for advantaged dictators, “Output focus” for disadvantaged dictators) will reduce the relative bias of advantaged dictators towards the advantaged recipients.

We combine the data from this experiment with a previous experiment in which dictators could freely choose what to look at. We will use regressions to evaluate the “difference in difference”.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Following standard Mouselab protocols, we will exclude information that was revealed for less than 200 ms.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will recruit 400 dictators from the Prolific platform. Dictators will be evenly split between the 4 between subject conditions (i.e. 100 in each cell). We recruit a corresponding number of receivers.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will investigate whether the impact of merit/output information on giving differs between advantaged and disadvantaged dictators.

We will correlate giving and attention with several additional elicitations in the questionnaire on perceptions of fairness.

D Instructions

The instructions for the dictators in the experiment are shown below, together with the comprehension questions. The instructions were presented in several decks of slides. Participants could move across slides clicking on two buttons on the sides of the screen. Comprehension questions were presented on a separate page. Participants could move back to the instruction from the page with the comprehension questions.

D.1 Day 1

For the recipients the third slide of this set read: "This study consists of two sessions. However, you will only participate in Session 1 that takes place today. Session 1 lasts about 34 minutes. The reward for completion is £2.85.

On top of the base earnings, in this study you can earn a **bonus.**

The amount of the bonus may depend on:

- Your **performance**
- Your **decisions** or the decisions of another participant
- **Luck**

The Ethics Committee Economics and Business (EBEC) of the University of Amsterdam has approved our study (EC 20200214090248). You can contact our Ethics Committee writing to secbs-abs@uva.nl.

To receive the approval we committed ***not*** to use misleading or untruthful instructions.

UNIVERSITEIT VAN AMSTERDAM

Please answer the questions on the next webpage.

Comprehension questions:

1. Your bonus might depend on the decisions taken by another participant. T/F [correct: T]
2. You can complete this study using a mobile device. T/F [F]
3. According to the ethical protocol under which we run this study, all the instructions you read must be truthful and not misleading. T/F [T]
4. You need to complete both sessions of this study for your submission to be approved. T/F [T]

Instructions for Session 1

In this session, you will do 8 tasks.

In each task you will have to answer several questions.

Within each task, every correct answer gives the same monetary reward.

However, **different tasks** give different monetary rewards per correct answer.

There are two possible reward levels, **high** and **low.**

We will split participants into two groups:

50% in the High Reward Group

50% in the Low Reward Group

What is the difference?

Participants in the **High Reward Group** will always receive the **High Reward** per correct answer.

Participants in the **Low Reward Group** will always receive the **Low Reward** per correct answer.

Luck decides if you are in the High or Low Reward Group.

The computer determines your group randomly. You have 50% chance to be in either group

If the computer assigns you to the **High Reward Group**, then you will be in the High Reward Group for the **whole study**.

Similarly, if the computer assigns you to the **Low Reward Group**, then you will be in the Low Reward Group for the **whole study**.

You will know if you are in the High or Low Reward Group only at a later stage.

In each task, you are matched with a different participant

After each task, the reward that both you and the other participant earned flows into a common account.

At a later stage, we will explain how the money in this account is paid out to the members of the pair.

Please answer the questions on the next webpage.

1. In this study you have to complete BLANK tasks. [8]

2. There are 3 groups of participants. T/F [F]

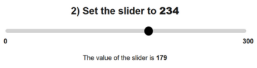
3. Luck determines if you are in the High Reward Group or in the Low Reward Group. T/F [T]

4. In some tasks, you will be in the High Reward Group, in others you will be in the Low Reward Group. T/F [T]

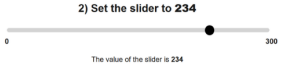
Tasks 1 and 2

You need to move a slider to a predetermined value.
You have **2 minutes** to move as many sliders as you can.

Incorrect slider



Correct slider



You get a reward per each slider you set to the correct value.

High Reward: £0.30 **Low Reward: £0.10**

Tasks 3 and 4

You need to answer a quiz.
You have **4 minutes** to answer as many questions as you can.

You get a reward per each correct answer.

High Reward: £0.90 **Low Reward: £0.30**

Tasks 5 and 6

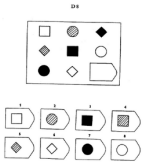
You need to count zeros in a table.
You have **3 minutes** to count the zeros in as many tables as you can.

You get a reward each time you correctly state the number of zeros.

High Reward: £0.30 **Low Reward: £0.10**

Tasks 7 and 8

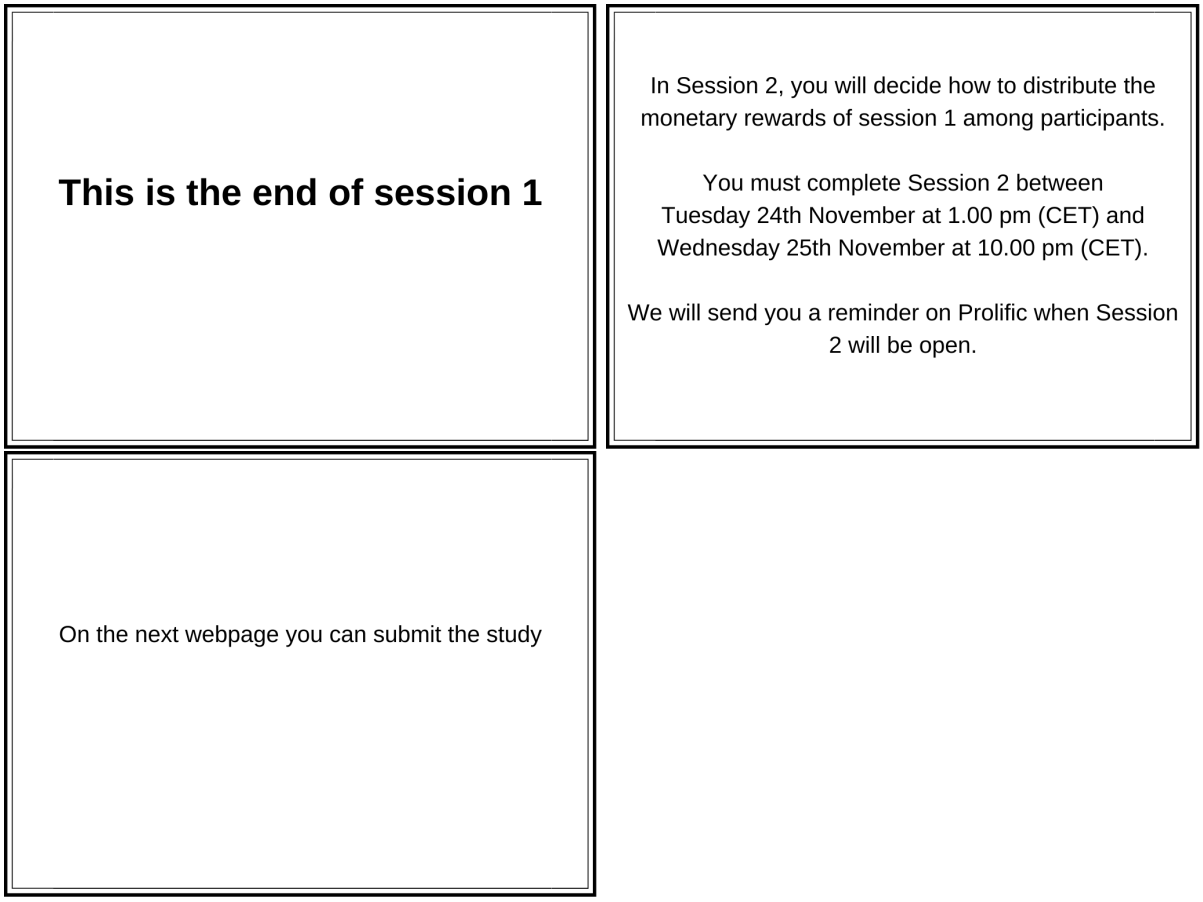
You need to find the missing tile in a figure.
You have **3 minutes** to complete as many figures as you can.



You get a reward each time you select the correct tile.

High Reward: £0.60 **Low Reward: £0.20**

The slides with the task instruction appeared before the relevant pair of tasks. To continue to the task, the participants had to correctly input the two possible pay-rates for the task.



In this last set of slides for the recipients, the second slide read: "We will approve your submission within two working days. The participants in Session 2 will decide how to distribute the monetary rewards of session 1 among participants. We will let you know your bonus and your pay rate with a Prolific message".

D.2 Day 2

Comprehension questions:

1. I confirm that I am using a laptop or desktop. Y/N [Yes]
2. Your performance on the tasks in Session 1 carries over into Session 2. T/F [True]
3. We commit to providing entirely accurate and truthful information in all aspects of this study. T/F [True]

Welcome!

This is a study conducted by researchers at the University of Amsterdam.

UNIVERSITEIT VAN AMSTERDAM

All the tasks in this study are **not mobile-compatible.**

You can only participate with a desk- or laptop, **preferably with a mouse** instead of a mousepad.

The Ethics Committee Economics and Business (EBEC) of the University of Amsterdam has approved our study. You can contact our Ethics Committee writing to secbs-abs@uva.nl.

To receive the approval we committed ***not*** to use misleading or untruthful instructions.

UNIVERSITEIT VAN AMSTERDAM

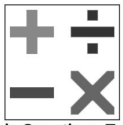
Welcome to Session 2 of this study.

This session consists of three parts that will last for approximately **55** minutes.

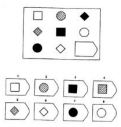
In Session 1, you completed a series of tasks online and you earned a reward for each correct answer.

Moving Sliders Task

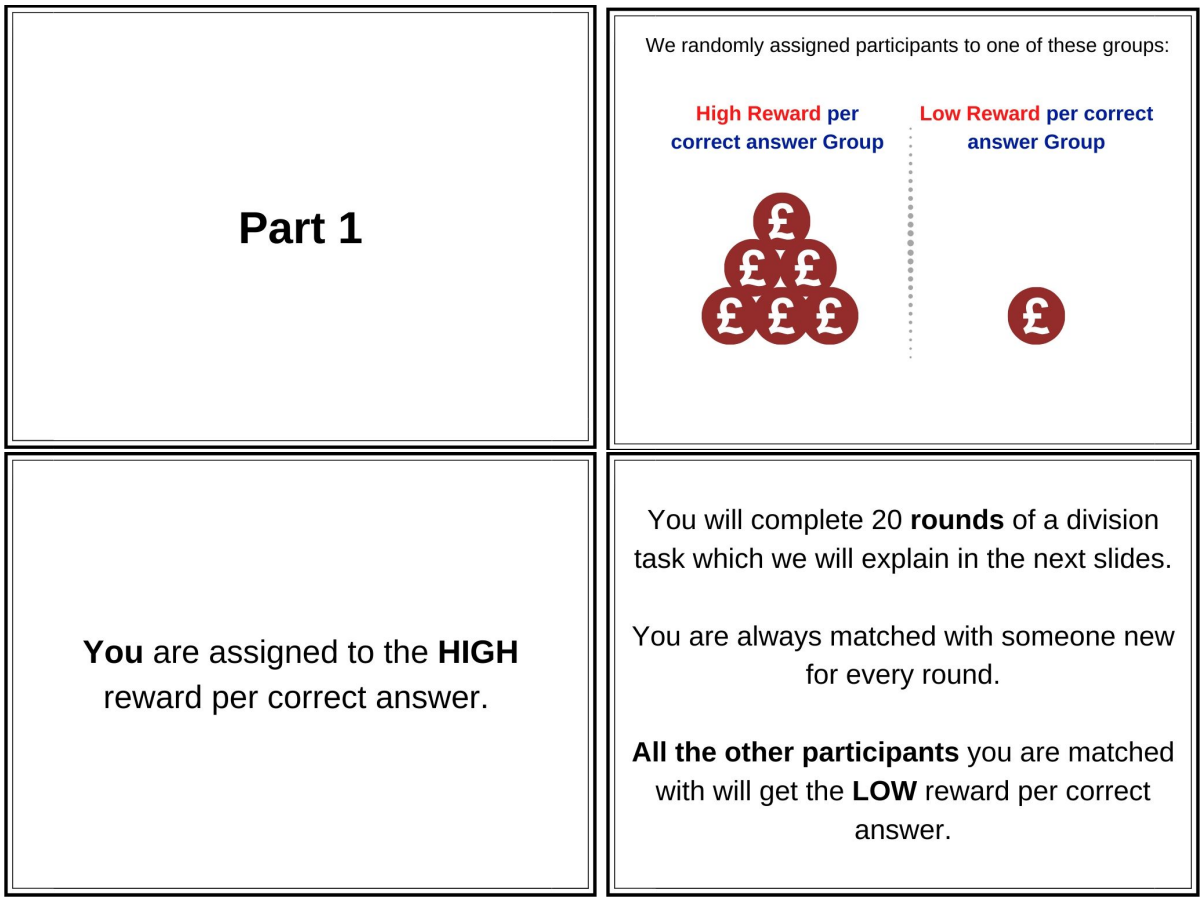
Counting Zeros Task



Quiz Questions Task



Missing Tiles Task



The second row shows the instructions for Advantaged participants, whereas Disadvantaged participants saw HIGH and LOW switched across slides. Disadvantaged participants were instructed that that they were assigned the LOW, and the other participants the HIGH, reward per correct answer.

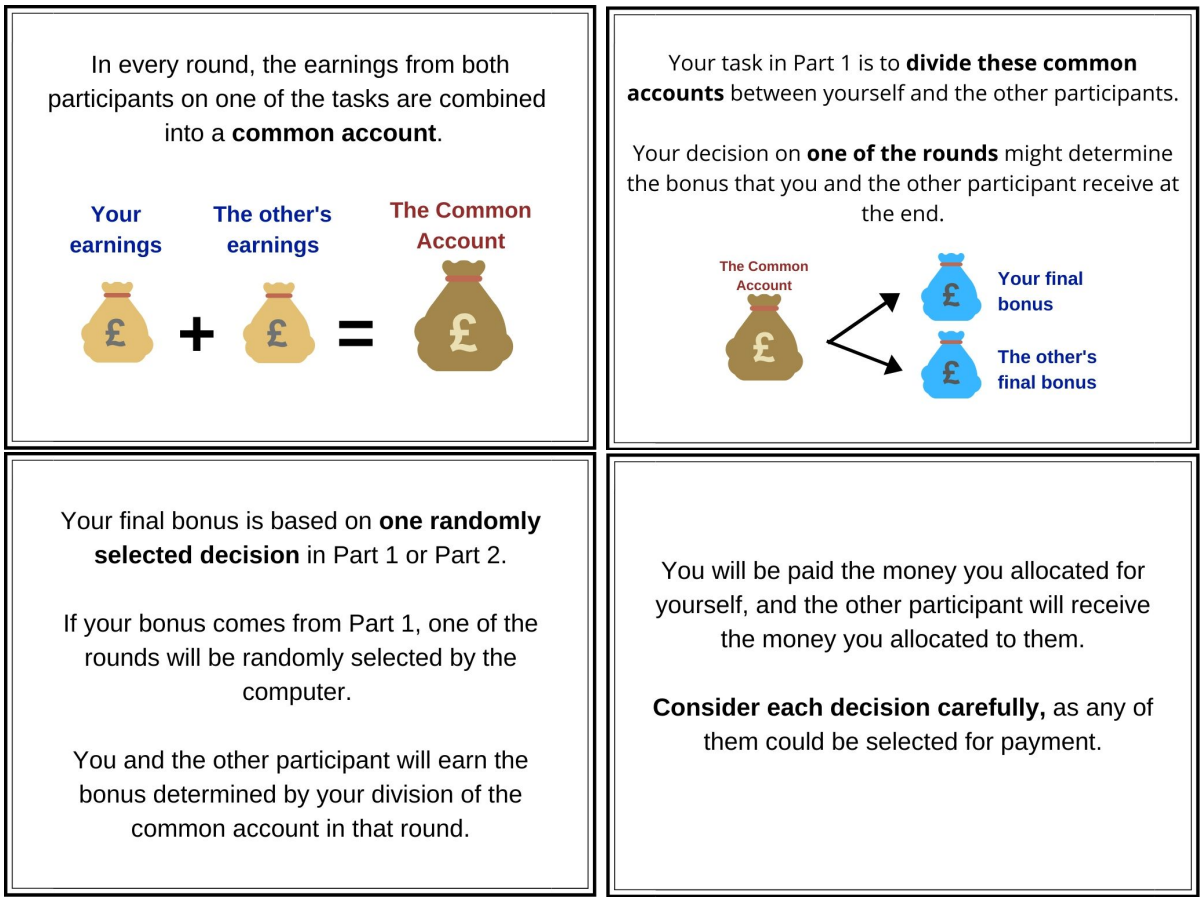
Comprehension questions:

1. Which reward condition were you and the other participants you are matched with assigned to? MULTIPLE CHOICE [Advantaged: You: High reward, Others: Low reward; Disadvantaged: You: Low reward, Others: High reward;]

2. What determines the common account on each round? MULTIPLE CHOICE [The combined amount you and the other participant earned on a single task from Session 1]

3. If Part 1 determines the bonus, how will you be paid? MULTIPLE CHOICE [The amount you gave yourself on a randomly selected round from Part 1]

4. If Part 1 determines the bonus, how will the other participant you are matched with be



paid? MULTIPLE CHOICE [The amount you gave them on a randomly selected round from Part 1]

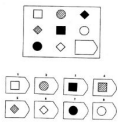
Before you decide on the division of the common account, we will show you some **additional information**.

First, you will see **which task** from Session 1 determined the common account for this round.

Moving Sliders Task

Counting Zeros Task

Quiz Questions Task



Missing TilesTask

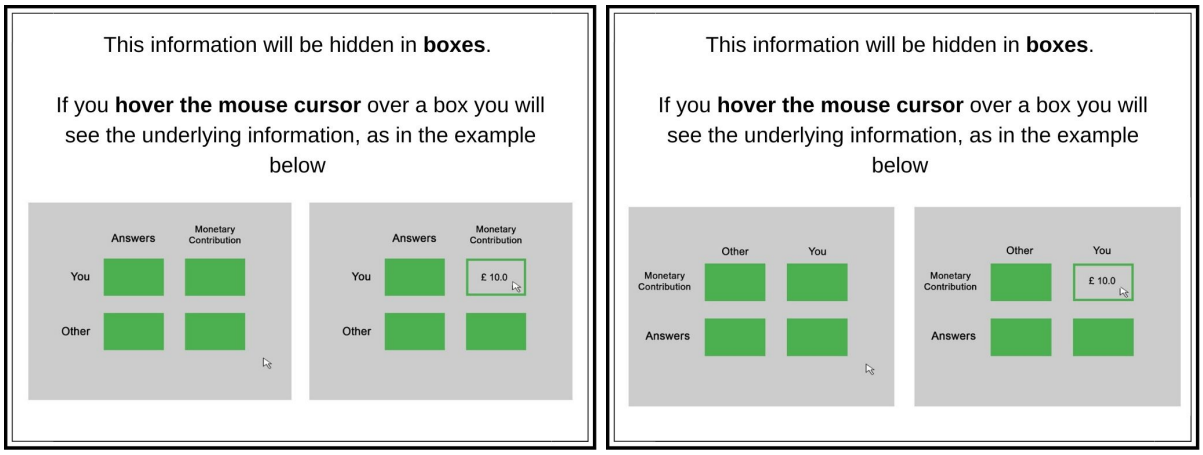
On the same page, you will also see the **total earnings** in the common account that you will divide, as in the example below:

Task: Counting Zeros

Common account: £10

Next, you have the chance to learn how both participants contributed to the common account. There are two kinds of information:

- **Correct answers**: the number of correct answers you and the other participant each got on the task.
- **Monetary contribution**: the earnings you and the other participant each contributed on the task. This depends on the correct answers and the reward (high or low):
  correct answers x *reward per correct answer.*



Two examples of different information orientations. We used all 8 possible orientations of participant and contribution information between subjects, evenly divided across subjects accounting for Advantaged status and Focus treatment. Each subject only saw one orientation to allow them to develop information-seeking patterns.

The information boxes are available for **6 seconds.**

Within this time limit, you can decide which and how many boxes to open. Boxes can be opened more than once.

At times, the program might close some boxes. If this happens, you can't open those boxes again in that round.

You will have the opportunity to practice with these boxes on the next page.

The information boxes are filled with a placeholder number.

In the actual rounds, the information will be based on your performance in Session 1.

This practice will not count toward your bonus.

In the first practice round,
you can familiarize yourself with the **layout** of the information boxes for as long as you want.

In the second practice round,
you will familiarize yourself with the **timing** of the information boxes.

They will be presented for **6 seconds**, like in the actual rounds.

In summary:

- In 20 rounds you will divide a common account earned by yourself and another participant.
- Each participants' contribution depends on the correct answers and the reward per correct answer on a single task.
- Before the division, you can inform yourself about correct answers and monetary contributions by hovering your cursor over information boxes.
- Any round could be chosen for payment at the end of the session.

In the first shown slide, the last paragraph "At times, the program might close some boxes" was only included in Experiment 2 and left out in Experiment 1.

Comprehension questions:

1. On the information screen, what does "correct answers" refer to? MULTIPLE CHOICE [The number of answers you and the other participant each got correct on that task]

2. On the information screen, what does "monetary contribution" refer to? MULTIPLE CHOICE [The earnings (correct answers X reward rate) you and the other participant each contributed to the common account on that task]

Comprehension questions:

1. Which reward condition were Player High and Player Low assigned to? MULTIPLE CHOICE [Player High: High reward, Player Low: Low reward]

2. If Part 2 determines the bonus, how will you be paid? MULTIPLE CHOICE [A set 1 pound bonus]

3. If Part 2 determines the bonus, how will Player High and Player Low be paid? MULTIPLE CHOICE [The amount you gave to each of them on a randomly selected round from Part 2]

Part 2

Thank you for completing Part 1 of the Session 2.

Now, we will proceed to Part 2 of this study.

In this part of the study, you will do the same division task as the previous part.

However, you will divide common accounts for **two other, anonymous players** instead:

Player High and Player Low

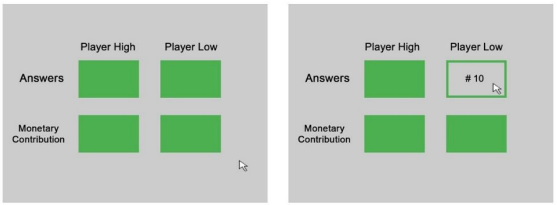
Just like you, both Player High and Player Low have also completed the series of tasks online in Session 1.

Player **High** receives a **HIGH Reward** per correct answer.
Player **Low** receives a **LOW Reward** per correct answer.

As in Part 1, you will be presented with information boxes with Player High's and Player Low's number of correct answers and earnings for **6 seconds**.

And you can **hover the mouse cursor over a box** to see the underlying information.

This is the information layout you will see



In row 2, the right slide switched the information about Players High and Low for Disadvantaged participants so Player Low was described first. The last slide showing the orientation of information varied based on the participant's information orientation. Here, the orientation matched that of Involved trials such that for Advantaged players, Player High's information was in the same row or column as Self information, and Player Low's information was in the same row or column as Self information for Disadvantaged players.

You will complete **20 rounds** of divisions.

For every round, you will make the decision for a different pair of players.

As a reminder:

Your bonus will be based on **randomly selected decision** from **either Part 1 or Part 2**.

If the computer selects one of your decisions from Part 2,

Player High and Player Low from a randomly selected round will be paid according to your decision.

You will receive a fixed bonus of £1.

Consider each decision carefully as any of them could be randomly selected for payment towards other participants at the end.

Social Norms

Part 3

In Part 3 we will ask you several questions.

At the end of the session, the computer will select one question from Part 3 at random.

In addition to your bonus from Part 1 or Part 2, you can win a bonus depending on your answer to this question.

We will give you more precise instructions about the bonus as you proceed through Part 3

Part 3.1

In the questions below, please give us your best estimate.

You will earn £1 if you are within 5% of the correct answer.

Elicitation questions

1. We selected a random task from Session 1 of the experiment and compared the task performance of 100 members of the HIGH group with the task performance of 100 members of the LOW group. The monetary earnings each person contributed is measured as the number of correct answers in the task times the reward rate. Remember that the reward rate per correct answer was higher in the HIGH group than in the LOW group.

   In how many of these 100 comparisons do you think that the member of the LOW group produced a larger monetary contribution than members of the HIGH group?

2. In Part 1, you were matched with 20 different participants and saw information on both your task performance and the task performance of the matched participants.

In how many of these 20 rounds did the participant you were matched with answered more question correctly than you did?[17]

[17]We asked these two questions only in Experiment 2 treatments.

Part 3.2

In the questions below, *morally appropriate* refers to an action that is "correct", "fair", or "ethical" according to your values and morality.

Elicitation questions How did you decide how to split the common account? [OPEN QUESTION]

According to your moral values, how would you judge the following ways of splitting the common account?[18]

1. Giving to each participant the monetary contribution he/she produced in Session 1
2. Giving an equal amount to each participant
3. Splitting the account considering only the number of correct answers of each participant in Session 1
4. Keeping all for oneself

[Possible answers: Very morally inappropriate, Somewhat morally inappropriate, Somewhat morally appropriate, Very morally appropriate]

[18]The order of the norms questions is randomized at the individual level and it is kept the same across the different elicitation screen. That is if a participant sees the questions in the order meritocratic, libertarian, egalitarian in the screen about the moral norms, this order is preserved in the following screens as well.

Part 3.3

You will now have to judge whether the behavior described in some statements is *socially appropriate*.

Those statements are the same you read in the previous webpage.

Socially appropriate behavior refers to an action that is "correct", "fair", or "ethical" according to most participants.

Bonus

If the computer selects one question from Part 3.3 for payment,

- We will check how participants that split the common account judged the social appropriateness of the behavior described in the question.
- You will win an additional £1 if your judgment coincides with the most common judgment.

For example, you will see a question like this:

1) How do you judge the following behavior?

"*I split the common account giving to each participant the monetary contribution he/she produced in Session 1.*"

- ○ Very socially inappropriate
- ○ Somewhat socially inappropriate
- ○ Somewhat socially appropriate
- ○ Very socially appropriate

You will win a £1 bonus if you pick the answer that is selected with the highest frequency by the other participants.

Please answer the questions on the next webpage.

Comprehension questions

1. For socially appropriate we mean an action that: MULTIPLE CHOICE [Cost people will find "correct", "fair", or "ethical"]

2. If a question from Part 3.2 is selected for payment, you earn a bonus of £BLANK if you: MULTIPLE CHOICE [pick the answer that is selected with the highest frequency by the other participants that divided the common account.]

Elicitation questions:

Are the following ways of splitting the common account socially appropriate? Remember to select the answer you think is most common.

1. Giving to each participant the monetary contribution he/she produced in Session 1

2. Giving an equal amount to each participant

3. Splitting the account considering only the number of correct answers of each participant in Session 1

[Possible answers: Very socially inappropriate, Somewhat socially inappropriate, Somewhat socially appropriate, Very socially appropriate]

Part 3.4

For the next questions you will have to guess how a group of participants judged some behavior.

The groups that you will have to consider are:

- participants that a) received a **low reward** per each correct answer in Session 1 and b) that split the common account in Session 2
- participants that a) received a **high reward** per each correct answer in Session 1 and b) that split the common account in Session 2

Bonus

As before, if the computer selects one question from Part 3.4 for payment:

- We will check which is the most common judgment among the group specified by the question.
- You will win an additional £1 if guessed what is the most common answer in that group.

For example, you will see a question like this:

Consider the group of participants that a) received a *low reward* per correct answer in Session 1 and b) split the common account in Session 2

1) How do you think most of participants in this group judged the statement below ?

"*I split the common account giving to each participant the monetary contribution he/she produced in Session 1.*"

- ○ Very socially inappropriate
- ○ Somewhat socially inappropriate
- ○ Somewhat socially appropriate
- ○ Very socially appropriate

You will win a £1 if you guess the most common judgment among the group described in the question.

Please answer the questions on the next webpage.

Comprehension questions In Part 3.3 you will have to guess the way most participants in some groups judged a statement. Among the groups below, tick all the ones you will have to consider.

- A group composed of participants that a) received a *low reward* per correct answer in Session 1 and b) split the common account in Session 2 [Correct]
- A group composed of participants that a) received a *high reward* per correct answer in Session 1 and b) split the common account in Session 2 [Correct]
- A group composed of participants that a) received a *low reward* per correct answer in Session 1 and b) *did not* split the common account in Session 2
- A group composed of participants that a) received a *high reward* per correct answer in Session 1 and b) *did not* split the common account in Session 2

Elicitation questions

Consider the group of participants that a) received a *HIGH REWARD* per correct answer in Session 1 and b) split the common account in Session 2

How do you think most of participants in this group judged the following ways of splitting the common account?

1. Giving to each participant the monetary contribution he/she produced in Session 1
2. Giving an equal amount to each participant
3. Splitting the account considering only the number of correct answers of each participant in Session 1

Now, consider the group of participants that a) received a *LOW REWARD* per correct answer in Session 1 and b) split the common account in Session 2

How do you think most of participants in this group judged the following ways of splitting the common account?

1. Giving to each participant the monetary contribution he/she produced in Session 1
2. Giving an equal amount to each participant

3. Splitting the account considering only the number of correct answers of each participant in Session 1

Questionnaire, page 1 **Please complete the following short survey.**

1. Age: [Open-ended question]
2. Gender: [Possible answer: Man, Woman, Other]
3. What is your nationality? [List of all countries in the World]
4. Generally speaking, where do you place yourself on the left-right political spectrum? [Possible answers: left, center-left, center, center-right, right]
5. How much do you agree with this statement? "The government should take measures to reduce differences in income levels." [Possible answers: Completely disagree, Somewhat disagree, Somewhat agree, Completely agree]
6. What is the highest level of school you have completed or the highest degree you have received? [Possible answers: Less than high school degree, High school degree, Some University but no degree, Bachelor degree, Master degree, Doctoral degree]
7. How much total combined money did all members of your HOUSEHOLD earn last year? [7 different income brackets]

We ask you the questions below to check whether we need to improve the study. As for every other question in the study, the approval of your submission does not depend on your answers.

8. Was there anything in the instructions that was unclear? [Open-ended question]
9. How attentive and focused were you in the last rounds of Part 1? [1 to 10 scale. 1 = Not at all attentive or focused, 10 = Completely attentive or focused]
10. How attentive and focused were you in the last rounds of Part 2? [1 to 10 scale. 1 = Not at all attentive or focused, 10 = Completely attentive or focused]
11. What do you think the aim of this study is? [Open-ended question]
12. Do you have any remark or suggestion? [Open-ended question]

Questionnaire, page 2 **Please answer the question below.**

In your experience, was/were there any box(es) that was/were more likely to be closed by the program?" [Open-ended question]

Questionnaire, page 3 **Please answer the following questions.**

How morally appropriate would you consider the following ways to use the information?

1. using exclusively information about the **participants' monetary contribution** in a task to decide how to split the common account? [1 to 5 scale. 1 = Very inappropriate, 5 = Very appropriate]

2. How morally appropriate would you consider using exclusively information about the **participants' number of correct answers** in a task to decide how to split the common account? [1 to 5 scale. 1 = Very inappropriate, 5 = Very appropriate]