Debiasing preferences over redistribution: An experiment[†]

Romain Espinosa[1], Bruno Deffains[2], and Christian Thöni[*3]

[1]CNRS, CREM - Université de Rennes 1, France
[2]CRED - Université Panthéon-Assas, Institut Universitaire de France, France
[3]University of Lausanne, Switzerland

February 24, 2020

Abstract

We study the manipulation of preferences over redistribution. Previous work showed that preferences over redistribution are malleable by the experience of success or failure in a preceding real-effort task. We manipulate the information subjects receive about the importance of chance relative to effort in determining success. We investigate the effect of this manipulation on (i) subjects' redistribution choices affecting third parties, and (ii) preferences for redistributive taxation. Our results show that informing the subjects about the relative importance of chance after the real-effort task does not mitigate the self-serving bias in redistribution choices. Only providing full information before the real-effort task prevents the emergence of the self-serving bias.

Keywords: Redistribution, Self-serving bias, Debiasing, Experiment

1 Introduction

Free market economies tend to result in unequal allocations of resources. While all modern societies put forward some form of intervention to mitigate the gap between rich and poor, redistributive measures are subject to intense political debates. Behind the Rawlsian veil of ignorance, all citizens might unanimously agree on a redistributive system. Real world debates are complicated by the fact that individuals know their position in the socio-economic hierarchy, which gives rise to opposing views about redistribution. Alesina and Giuliano (2011) show that preferences over redistribution depend on personal characteristics such as gender, race and socioeconomic status. Given that usually the poor outnumber the rich by far, it is a puzzle why democratic systems do not result in more extensive redistribution. Bonica et al. (2013) explore this question with respect to the U.S. and note that one of the explanations might be increased polarization between the Democrats and the Republicans, which makes it more difficult to find a compromise.[1]

While increasing inequality might be an important cause of the polarization on preferences for redistribution, Deffains et al. (2016) recently argued that an additional force is at play: Building on psychological research on self-serving biases, they show experimentally that the experience of success or failure shifts individuals' views on redistribution, even in the absence of any selfish motives. In their experiment, subjects are performing a real-effort task for which

---

[*]Corresponding author, e-mail: christian.thoeni@unil.ch.

[1]Weinzierl (2017) provides survey evidence showing that respondents do not seek full equalization of unequal outcomes produced entirely by randomness.

they are randomly assigned to either a success or a failure condition. After this, subjects are asked to redistribute wealth from a rich to a poor subject, a task that leaves their own payoff unaffected. The design deliberately introduces ambiguity as to whether success or failure is due to internal factors (insufficient effort or ability), or due to external factors (the nature of the task). The results show that subjects who had been successful in the real-effort task redistribute less than their less successful counterparts do. In addition, a questionnaire reveals that successful subjects tend to attribute their success to internal factors, while unsuccessful subjects attribute their failure to external factors. This 'political self-serving bias' contributes to the polarization of the views on redistribution: Not only do those who are born with a golden spoon in their mouths oppose redistributive measures because they are net payers, but also because they mistakenly think they earned their position in the hierarchy.

The polarization of views on redistribution documented by Deffains et al. (2016) is presumably due to the high level of ambiguity with respect to the relative importance of internal and external factors. To what extent the degree of ambiguity in the experiment approximates the real world is obviously difficult to assess. While the importance of intergenerational transmission of economic status is part of an academic debate (e.g. Bowles and Gintis, 2002), it seems not unlikely that an overestimation of the internal factors among relatively successful individuals is also pertinent outside the laboratory. More information about the relative importance of external and internal factors might thus reduce the polarization and facilitate reaching a consensus in the political process.

This study builds on the method of Deffains et al. (2016) and investigates to which extent the political self-serving bias is stable to adding information about the relative importance of external factors. Like Deffains et al. (2016) we elicit behavior in (i) a disinterested dictator game, where participants redistribute money among two other participants, and (ii) a game in which subjects choose redistributive taxation affecting their own future payoffs. For both dependent variables we confirm the earlier findings and show that—when there is high ambiguity in terms of the relative importance of external and internal factors—a political self-serving bias arises. We then investigate the stability of the bias. To do so we provide information about the relative importance of the external factors. In particular, we inform the participants that their success or failure in the real-effort task was almost fully determined by external factors. We provide this information either before or after the realization of success or failure. We find that the self-serving bias survives if the information is provided after the realization of success or failure. If, however, we provide the information that the deck is stacked before the realization of success or failure, then there is no systematic change in redistributive preferences.

The rest of the paper is organized as follows. In Section 2, we discuss the related literature, Section 3 describes the experiment and the hypotheses. In Section 4 and 5 we discuss the results and Section 6 concludes.

2 Theoretical and experimental background

Preferences for redistribution have received great attention in the experimental literature. A first line or research investigated the role of self-interest (Hoffman and Spitzer, 1985; Bjerk, 2016; Gerber et al., 2019), social preferences (Tyran and Sausgruber, 2006; Ackert et al., 2007; Balafoutas et al., 2013), or social identity (Klor and Shayo, 2010) for the determination of redistributive systems. Durante et al. (2014) provide estimates for the relative importance of selfish and fairness motives. Kittel et al. (2017) build on their framework and add the role of power as a determinant for redistributive preferences. Eisenkopf et al. (2013) study the effect of unequal access to education on the demand for redistribution.

In a second strand of studies, economists have sought to determine how individuals are willing to redistribute under uncertainty about their position in the income distribution (Schildberg-Hörisch, 2010). In a survey study, Karadja et al. (2017) informed respondents about their relative

position in the income distribution and show that respondents who are richer than they initially thought demand less redistribution. A series of studies have investigated how preferences for redistribution are affected by economic experience (Frohlich and Oppenheimer, 1990; Cabrales et al., 2012; Cappelen et al., 2007; Großer and Reuben, 2013).

One of the main discussions developed in this literature concerns the relative importance of luck and effort in the determination of individual outcomes (Kataria and Montinari, 2012; Cappelen and Tungodden, 2017).[2] Konow (2000) shows that individuals act according to the *accountability principle*, i.e., they use redistribution systems to compensate (bad) luck and to reward efforts. Evidence suggests, however, that individuals tend to deviate from the accountability principle when they have interests at stake: Babcock et al. (1995) and Babcock and Loewenstein (1997) show that people tend to exhibit a self-serving bias, i.e., their view about fairness is distorted to their advantage (see also Rodriguez-Lara and Moreno-Garrido, 2012; Ubeda, 2014; Gino et al., 2016). Deffains et al. (2016) demonstrate the existence of a political self-serving bias: when individuals succeed (fail) in a task, they are more (less) likely to attribute the outcome to their effort rather than luck. Consequently, these distorted attributions lead to diverging redistributive preferences.

Our experimental design is also related to a strand of literature studying situations in which subjects' decisions do not affect their own payoff, i.e., they act in the role of an impartial spectator (Croson and Konow, 2009; Konow, 2009; Tinghög et al., 2017; Konow et al., 2019; Luhan et al., 2019; Cassar and Klein, 2019).

The demonstration of the existence of self-serving biases for redistribution raises the question of how to mitigate the gap between biased individuals. Larrick (2004) distinguishes two main techniques for debiasing individuals in general: using one bias to offset another (i.e., rebiasing), or changing the decision environment. The effectiveness of these techniques has been shown to vary with the type of bias under consideration. Arkes (1991) discriminates between three types of judgment errors: strategy-based judgment errors (subjects take suboptimal decisions because elaborating more subtle strategies is too costly), association-based judgment errors (misperception of the probabilities of the states of the world), and psychophysically based errors (like e.g. sunk cost effects or framing effects). Strategy-based judgment errors are usually corrected by incentivizing individuals to develop more elaborated strategies (e.g., increasing the rewards associated to correct reactions). Psychophysically based errors can be compensated by shifting the reference point considered by individuals in their S-shaped utility function.

Remedies to mitigate association-based judgment errors like the self-serving bias mainly consist in three types of strategies: rebiasing, forcing participants to consider alternative reasoning, and emphasizing elements necessary to form correct beliefs. First, rebiasing is a relatively unpopular and risky strategy. Rebiasing requires indeed a good assessment of the pre-existing bias and a good calibration of the counter-bias. Should one of these two elements fail, rebiasing can either overshoot and generate an inverted bias, or fall short and fail at resorbing the bias. For instance, Jolls and Sunstein (2006) propose to use availability heuristics to correct for optimistic beliefs of consumers: by telling them anecdotes, one can increase their perceived probability of some states of the world. Second, forcing participants to consider alternative reasoning has been shown to mitigate biases (Slovic and Fischhoff, 1977; Koriat et al., 1980; Anderson, 1982, 1983; Lord et al., 1984; Babcock et al., 1997). A third approach proposes to make the information necessary for correct inference more salient (e.g., Nisbett et al., 1983): the objective is to emphasize useful information in order to increase the likelihood that the information is taken into account when individuals form their beliefs. In our experiment, we focus on this third debiasing strategy and provide subjects with specific information about the relative importance of internal and external factors.

---

[2]A related strand of literature studies similar effects in bargaining (Gächter and Riedl, 2005; Karagözoğlu and Riedl, 2015), dictator games (Erkal et al., 2011; Spiekermann and Weiss, 2016), money burning (Fehr, 2018), and effort provision (Kovárík et al., 2018).

3 Experimental Design

Our experimental protocol starts with inducing the experience of success or failure in a real-effort task. We define success (failure) as performing above (below) the median performance in the real-effort task. In our design, success and failure is almost exclusively determined by a random draw, assigning subjects to either a hard or an easy task. We conduct three treatments, varying in the degree to which subjects are informed about the relative importance of luck and effort in determining success and failure. In the *Baseline* treatment, we replicate Deffains et al. (2016) and inform subjects only about the existence of the hard and easy task and the random assignment procedure. We do not inform the subjects whether they were assigned to the hard or easy task. In the other treatments, we inform the subjects about the task they were assigned to and the relative importance of the random task assignment in determining success and failure. We conducted two treatments, differing with respect to the timing of the information. In treatment *PostTask*, subjects learn after the realization of success and failure that the random task assignment was almost fully determining the outcome. In treatment *PreTask*, subjects receive this information prior to the real-effort task. Subsequently we elicit the main dependent measure—redistributive preferences—in two games, the *Disinterested Dictator Game* and the *Redistributive Taxation Game*. In the following, we describe the elements of the experimental protocol in more detail.

3.1 Real-Effort Task

In the real-effort task, subjects were presented with sequences of binary digits, with four to thirteen digits each. Subjects had to count and indicate the number of ones in each sequence. There were five consecutive screens with a time limit of 30 seconds per screen. Correct answers were rewarded. In each session, half of the subjects were randomly assigned to the 'hard task', the other half to the 'easy task'. In the hard task, screens were made of 25 lines of binary digits. In the easy task, the last six lines were not displayed, and counted as correctly solved. The theoretical maximum number of tokens was identical in both tasks, but the tasks were designed such that it was very unlikely that a subject with the hard task would earn more tokens than a subject with the easy task. After completion of the five screens, we used the total number of tokens earned in the real-effort task to perform a median split of the subjects within a session. Subjects who earned more tokens than the median were told that they performed above median. The wording on the screen was:

> Congratulations! You performed better than most of the participants!

For the subjects with below median performance the message read:

> What a pity! You performed worse than most of the participants!

In the following, we label above median performers as *overachievers*, and below median performers as *underachievers*. Participants were told that the maximum possible earnings were the same in both tasks. While the two tasks clearly varied in difficulty, even the easy task was designed such that none of the participants managed to solve it perfectly, given the time limit. In addition, participants could not observe other participants' tasks. Consequently, in absence of information about the assigned tasks, participants were unable to deduce which task they were assigned to. Information about the task assignment was our treatment variation, which we will describe next.

3.2 Treatments

We conducted three treatments differing in the amount and the timing of information about the degree to which subjects' relative performance was determined by internal or external factors:

- *Baseline*: This treatment replicates the procedures by Deffains et al. (2016), where no additional information was provided. Note that in all our treatments participants are aware of the fact that (i) there is a hard and an easy task, and (ii) there is random assignment to task with equal probabilities.

- *PostTask*: In this treatment, participants receive additional information after the real-effort task and the realization of under-/overachiever. First, we informed subjects about the task they were assigned to (hard/easy). Second, we explained that in our previous sessions all participants with the hard task ended up as underachievers and vice versa. We based this information on the perfect separation observed in previous experimental research using the same real-effort task (Deffains et al., 2016). Finally, we told participants that the game was designed such that it was "virtually impossible" to end up as overachiever (resp. underachiever) when given the hard (resp. easy) task.[3]

- *PreTask*: In this treatment we gave participants the same information as in *PostTask*, but participants received the information *prior* to the real-effort task. They were told that the game was designed such as to produce full separation. After the real-effort task, when participants were informed whether they performed better or worse than the median, we also told them the nature of the task they received.

3.3 Disinterested Dictator Game

For the *Disinterested Dictator Game* (DDG) we selected two participants (the 'targets') randomly among all participants of the session. The remaining participants (the 'disinterested dictators') had then the possibility to redistribute tokens from the wealthier to the poorer target. To avoid comparison, disinterested dictators were not informed about the absolute level of the two targets' payoffs. Instead, they could reallocate the difference between the wealthier and the poorer subject's payoffs from the former to the latter. In particular, they had to indicate a percentage of redistribution. Figure 1 shows an excerpt of the experimental interface. When subjects entered the stage, they were shown the two bars on the left indicating that the two targets' payoffs differ, but not allowing any inference about the absolute payoffs. Subjects select a redistribution level on a 12-point equidistant scale reaching from 0% (no redistribution) to 100% (transfer all the money at disposal from the wealthier target to the poorer target).[4] Upon choosing a redistribution percentage, the two bars on the right appeared as a visual confirmation of the selected action (Figure 1 shows the case of 45.5% redistribution). Subjects could revise their decision and get an updated bar chart until they were satisfied with the result.

All participants were told that the decision of one disinterested dictator in the session would be randomly selected and implemented. Participants were also explicitly told that redistribution would concern the two targets only, and that all other subjects would not be affected by any redistribution mechanism in this task. Prior to the decision, disinterested dictators were reminded that targets may have faced different tasks. After all disinterested dictators made their choice, one redistribution proposal was randomly selected and implemented. Payoff information from the first game was withheld until the end of the experiment.[5]

---

[3]In few sessions of the *PostTask* treatment, participants were not given this last piece of information, namely that the game was designed such as randomness almost completely determined success. We found no statistical difference neither for the demand of redistribution in the DDG ($p = .806$, Wilcoxon rank-sum test), nor in the RTG ($p = .652$), and we pool these observations for the remainder of the paper.

[4]The associated redistribution choices were: 0%, 9.1%, 18.2%, 27.3%, 36.4%, 45.5%, 54.5%, 63.6%, 72.7%, 81.8%, 90.9%, 100%. We deliberately excluded the 50-50 split (as e.g. in Bellemare et al., 2008).

[5]We slightly depart from the DDG introduced in Deffains et al. (2016) in three ways. First, we do not display the difference of tokens between the two targets, but let dictators choose on a percentage of redistribution. This ensures that all dictators have the same strategy space. Second, we make the consequences of the dictator's decision highly salient by means of the graphical representation. Third, participants do not learn their payoff at the end of the DGG, which minimizes spillovers on the subsequent game.

Figure 1: Screenshot of the Disinterested Dictator Game game.
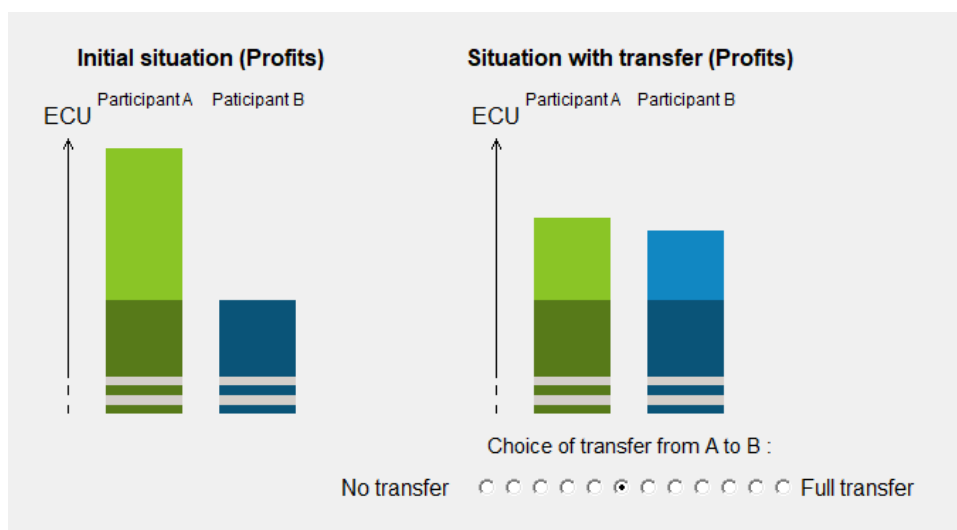
3.4 Redistributive Taxation Game

After the DDG, participants were given a new set of instructions for the *Redistributive Taxation Game* (RTG). In this game, subjects decided upon redistributive taxation for income to be earned in a real-effort task later in the experiment. Participants were told that they were going to perform a series of real-effort tasks different from the one before.[6] They were told that they could earn up to 80 tokens for the whole series of tasks, and that their final income would be subject to redistribution within a group of four participants. Before engaging in the real-effort task, participants in a group collectively decided on the redistributive rule for their group.

The redistribution rule was a simple redistributive tax scheme among the four participants in the group. Subjects could opt for a tax rate on their future real-effort task earnings from 0% to 100% by increments of 10 percentage points. All taxed revenue was redistributed in equal parts to the four participants. Consequently, with a tax of zero all subjects keep their earnings, while a tax of 100% means that all subjects in the group of four earn the same. Before taking a decision, participants went through extensive control questions with hypothetical situations to ensure that redistribution rules were well understood. (See Figures D1 to D4 in Appendix D for the screenshots.)

The decision about the redistributive tax was chosen by a random dictator mechanism. All subjects indicated their preferred tax rate, and in each group the decision of one randomly drawn subject was implemented.[7]

---

[6]The second series of tasks consisted in memorizing pictures and short video clips. After displaying each picture or video clip for about one minute, subjects had to answer questions regarding their contents. For this second real-effort task there was no hard/easy manipulation and no median split of the participants according to performance. The sole purpose of the second real-effort task was to generate variance in the earnings among subjects in a group to render the choice over redistribution in the RTG meaningful.

[7]After the random dictator mechanism participants were asked to bargain within their group of four to reach a consensus on the redistribution rule. The bargaining process consisted of five stages, in each of which participants could publicly announce their desired level of redistribution, and bargaining would stop in case of unanimous agreement on a tax rate. Both the random dictator and the result of the bargaining were potentially relevant for the outcome of the RTG. With 50% probability a group's final tax rate was determined by the random dictator rule; with the remaining probability it was determined by the bargaining. In the following we will focus on the private decision, as this is not confounded with strategic concerns.

3.5 Hypotheses

Previous findings show that the ambiguity in the *Baseline* protocol can induce a self-serving bias, which affects the demand and the supply of redistribution (Deffains et al., 2016). Our goal is to replicate this finding and to investigate whether adding information about the relative importance of luck in determining success and failure can serve as a debiasing device, such that differences in the redistribution decisions between the two groups disappear.

First, we expect our protocol to generate a self-serving bias among participants when causal attributions are unclear. We expect over- and underachievers to have different views about the extent to which internal and external factors affect the outcome when they lack additional information. Previous research showed that, when participants have diverging views about causal attributions, they also diverge in their redistribution preferences. The accountability principle discussed by Konow (2000) implies that participants tend to reward effort and to compensate for bad luck. While Konow (2000) provides evidence at the micro-level, the results by Alesina and Angeletos (2005) suggest that similar mechanisms are at play at the macro-level.

In our experiment, the self-serving bias predicts that succeeding (failing) in the task will increase one's propensity to believe that success is driven by effort (luck). Konow's accountability principle postulates that in redistribution choices react to the degree to which an observed distribution is caused by luck or effort. Combining the accountability principle with the self-serving bias, we predict that overachievers (underachievers) are less (more) supportive of redistribution, because they perceive the outcome to the real-effort task to result mainly from effort (luck). This prediction was confirmed in Deffains et al. (2016) and we use our *Baseline* condition to test whether their results replicate. Consequently, in this treatment we expect overachievers to redistribute less in the DDG and to opt for lower tax rates in the RTG.

The self-serving bias emerges in *Baseline* because causal attributions are unclear. In the two main treatments, we provide information about the mechanism leading to success or failure. Informing the participants about the mechanisms that determine the outcome should reduce the differences in the causal attributions across groups. Following Konow's accountability principle, groups with similar beliefs on the role of luck and effort should then display similar redistribution choices. In both *PostTask* and *PreTask*, causal attributions are extensively explained to the participants when they learn that the success and failure in the real-effort task is almost purely randomly determined. This leaves little room for diverging views on causal attributions. If participants process this information, overachievers and underachievers should have similar beliefs and should therefore choose similar levels of redistribution, both in the DDG and the RTG.

Previous works in social psychology (e.g., Guenther and Alicke 2008) suggest, however, that informing participants about causal attributions after they have formed their beliefs may not have the desired effect. The concept of *belief perseverance* states that people often tend to cling to their convictions when presented with contradictory information. In a seminal work, Ross et al. (1975) showed that participants who received positive feedback about their performance in a task were more likely to keep giving credit to the task than those who received negative feedback, even when they were told that the feedback was fabricated. According to the authors, perseverance takes place because individuals spontaneously form beliefs to explain the original feedback, and later they are reluctant to revise their beliefs in light of new information. In our *PostTask* treatment participants experience success or failure and, similar to the design of Ross et al. (1975), receive information about the mechanism ex-post. If belief perseverance is at play, we should expect similar results as in *Baseline*, i.e., the political self-serving bias, once established, might be immune to information about the causal mechanism.

Finally, in the treatment *PreTask* we inform subjects before entering the real-effort stage. This is arguably the strongest debiasing intervention, and we hypothesize that this treatment prevents the emergence of the self-serving bias, and, consequently produces similar levels of redistribution in both the DDG and the RTG.

4 Results

We ran seventeen sessions with 24 participants. In each session, two subjects were selected as targets for the DDG game and did not take a decision in this game. For the analysis, we dropped these observations from the sample, which leaves us with a total of 374 subjects. All sessions were run at the University of Strasbourg (January and June 2017). The sessions lasted about 45 min, and participants earned on average 13.7 euro. Seven sessions were devoted to the *Baseline* treatment, seven to the *PostTask* treatment, and three to the *PreTask* treatment. We start presenting the results with the real-effort task, followed by the Disinterested Dictator Game and the Redistribution System Game.

4.1 Real-effort task

The experiment starts with a real-effort task, in which subjects are asked to count the number of ones in binary sequences. This provides us with a measure for individual performance. Figure 2 shows the average performance of subjects in the two tasks. Subjects assigned to the easy task earned substantially more tokens than those in the hard task (16.5 vs. 10.3). The standard error spikes in Figure 2 indicate that the difference is highly significant ($p < 0.001$).

Figure 2: Performance in the *hard* and *easy* real-effort tasks.
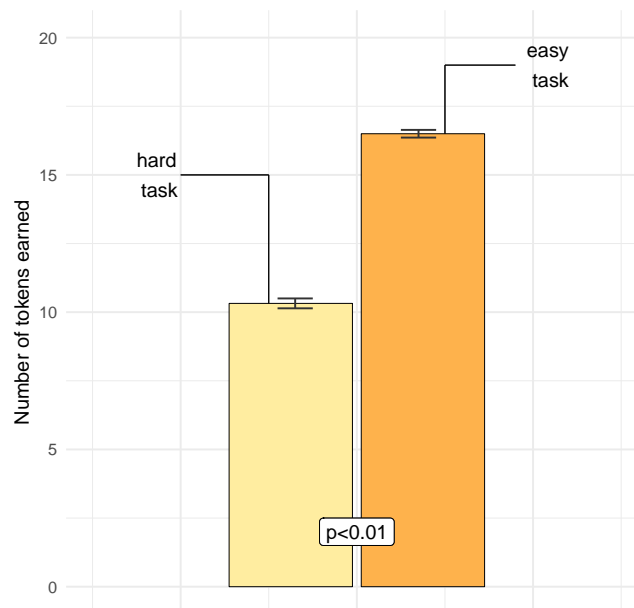
In the treatments *Baseline* and *PostTask* subjects are in the identical situation when commencing the real-effort task, while in *PreTask* subjects were informed that the task was designed to achieve separation. This information might have affected their motivation to exert effort. We use OLS regressions to test whether the performance in the real-effort task reacts to the treatment variations. We find very similar results in all treatments (see Table A2 in the Appendix).

Based on the performance measure we classify our subjects into overachievers (above median performance), and underachievers. While the two tasks produced widely different average outcomes, they were not fully separating. Six subjects in the hard task managed to become overachievers, and, by construction, six subjects in the easy task became underachievers. These switchers account for 3% of our sample, and, while this number seems too low to influence our result, we will account for the switchers in our analysis in various ways.[8]

---

[8] The fact that we observe switchers raises the question whether our design was deceptive, because we told

Our protocol induced the same level of information for both underachievers and overachievers regardless of their original task. It follows that participants were not able to deduce without further information whether they were assigned to the hard or easy task, such that presumably only the labelling as 'above the median' or 'below the median' affected their attributions. Consequently, the self-serving bias (SSB) can occur irrespective of the original task a participant was assigned to.

Figure 3: Histograms of the redistribution decisions in the DDG.
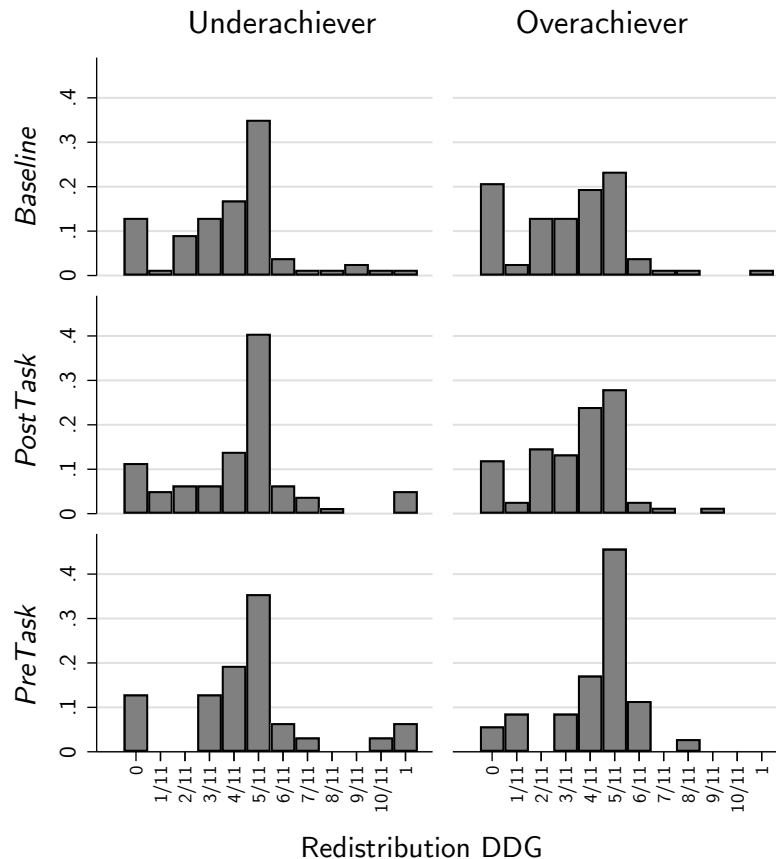

4.2 Disinterested Dictator Game

On average, subjects transfer 34.9% of the disposable tokens from the wealthier to the poorer target and most subjects redistribute at least some of the income. Figure 3 shows histograms of the redistribution decisions (by treatment and for under-and overachievers). In all cases, the most frequently chosen level is the option that almost equalizes the profits, but still leaves slightly more to the target who is initially wealthier (5/11). However, the histograms suggest that the tendency towards this option is stronger among underachievers than among overachievers in the treatments *Baseline* and *PostTask*, but not in *PreTask*.

Figure 4 shows the average redistribution percentage by over- and underachievers in the three treatments. In *Baseline*, we replicate the results of Deffains et al. (2016) and observe that

---

subjects that would be virtually impossible to switch. At the time we designed the experiment we based the wording of the instructions on observations from four sessions of our previous work (Deffains et al., 2016), where we used the same real effort task and observed perfect separation. We decided to keep the procedures constant between sessions despite the fact that we observed switchers in the experiments for the present study. In addition, the instructions read *virtually* impossible, which technically still concedes the possibility of switching.

Figure 4: Redistribution in the *Disinterested Dictator Game*.
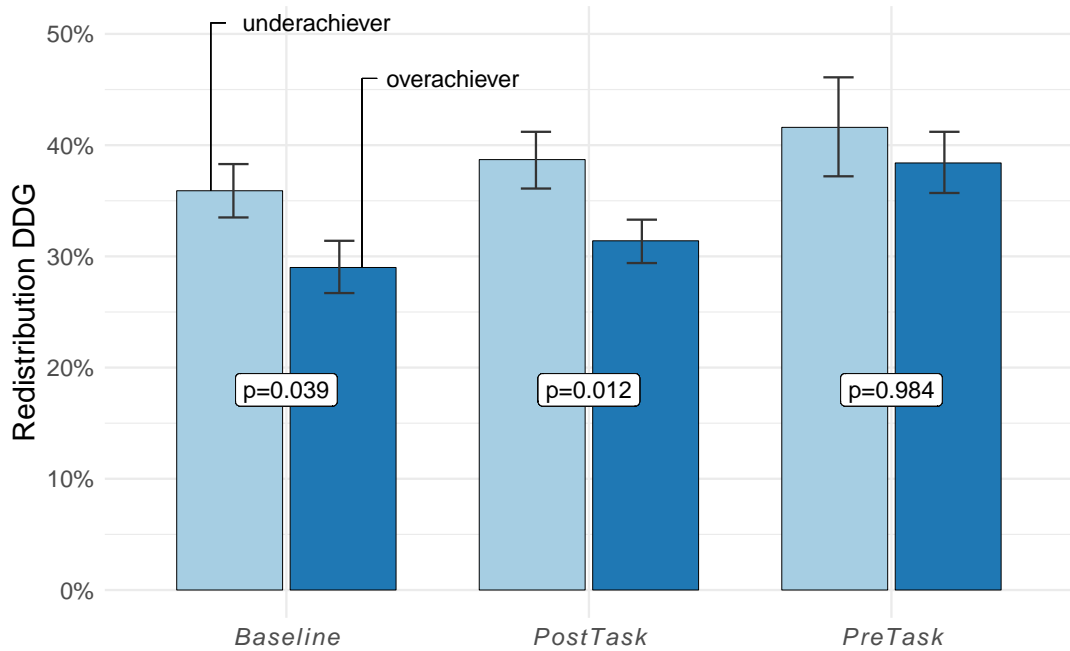
underachievers redistribute more than overachievers. On average, underachievers redistribute 35.9% compared to 29.0% for overachievers ($p = 0.039$, Wilcoxon rank-sum test).

Result 1. *In absence of information, underachievers redistribute significantly more from the rich to the poor target than overachievers do.*

Second, we investigate whether the gap between overachievers and underachievers disappears when we provide information about the relative importance of external and internal factors (henceforth the luck/effort ratio). Figure 4 provides mixed evidence. On one hand, we observe a similar difference in redistribution choices when the information is provided *after* subjects work on the real-effort task (treatment *PostTask*). In this case, underachievers propose to redistribute 38.7% on average, while overachievers prefer to redistribute only 31.4% ($p = 0.012$, Wilcoxon rank-sum test). On the other hand, the difference between overachievers and underachievers shrinks and becomes nonsignificant ($p = 0.984$, Wilcoxon rank-sum test) in the treatment *Pre-Task*. This suggests that information prior to the real-effort task weakens the emergence of the self-serving bias.

Result 2. *Underachievers and overachievers display similar redistribution preferences when information about the luck/effort ratio is provided prior to performing the real-effort task. When the information is provided ex post, overachievers redistribute significantly less from the rich to the poor target than underachievers do.*

Results 1 and 2 are supported by multivariate analyses presented in Table 1. We estimate the marginal effect of being an overachiever in each treatment on the redistribution decision in the first game for the 374 disinterested dictators. Given that participants chose from a set of discrete alternatives, we use ordered probit estimations. We report the estimated coefficients in the table together with the marginal effects associated with the overachiever status. Given the complexity of interpreting interaction coefficients in non-linear models (Ai and Norton, 2003; Norton et al., 2004), we report the marginal effects in percentage points in the table and detail our procedure to retrieve them in Appendix B. Specification (1) extends the results of the non-parametric tests presented above to the non-linear model. Specification (2) confirms the robustness of the results

Table 1: Regressions of the demand for redistribution in the DDG.

Variable	Redistribution DDG				
	(1)	(2)	(3)	(4)	(5)
PostTask	0.183	0.147	0.146	0.184	0.128
	(0.166)	(0.169)	(0.169)	(0.171)	(0.174)
PreTask	0.269	0.192	0.204	0.212	0.229
	(0.221)	(0.229)	(0.229)	(0.233)	(0.240)
Overachiever × *Baseline*	-0.350**	-0.344**	-0.357**	-0.344**	-0.406**
	(0.168)	(0.169)	(0.169)	(0.171)	(0.179)
Overachiever × *PostTask*	-0.417**	-0.410**	-0.424**	-0.420**	-0.434**
	(0.168)	(0.169)	(0.169)	(0.171)	(0.177)
Overachiever × *PreTask*	-0.0478	-0.0332	-0.0656	-0.0990	-0.166
	(0.255)	(0.257)	(0.260)	(0.264)	(0.284)
Age		0.00790	0.00898	0.00770	0.00806
		(0.0209)	(0.0211)	(0.0223)	(0.0209)
Political orientation		-0.0637**	-0.0658***	-0.0668**	-0.0642**
		(0.0253)	(0.0254)	(0.0260)	(0.0253)
Female		0.0298	0.0206	0.0123	0.0216
		(0.109)	(0.110)	(0.111)	(0.109)
Standardized performance	-0.165***	-0.155***	-0.170***	-0.170***	-0.179***
	(0.0563)	(0.0569)	(0.0595)	(0.0596)	(0.0590)
ME Overachiever in *Baseline*	-6.625**	-6.394**	-6.630**	-6.386**	-7.556**
	(3.166)	(3.126)	(3.130)	(3.162)	(3.328)
ME Overachiever in *PostTask*	-7.976**	-7.663**	-7.900**	-7.875**	-8.116**
	(3.192)	(3.136)	(3.142)	(3.183)	(3.304)
ME Overachiever in *PreTask*	-2.366	-1.578	-2.197	-2.796	-4.069
	(5.183)	(5.208)	(5.184)	(5.244)	(5.561)
Switchers	.	.	Dummy	Excluded	IV
Observations	374	374	374	362	374
LL	-716.84	-709.24	-708.70	-690.58	-750.28

Standard errors in parentheses
Ordered probit estimates, and marginal effects (ME) in percentage points.
*** p<0.01, ** p<0.05, * p<0.1

Figure 5: Tax rate in the *Redistributive Taxation Game*.
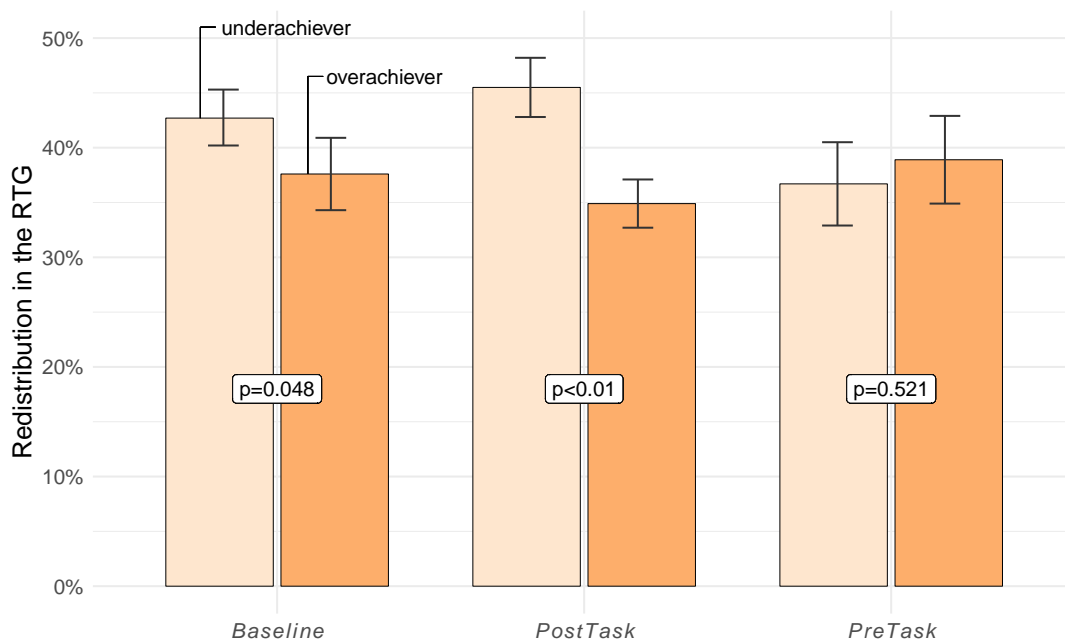
when controlling for age, gender, political orientation, and standardized performance[9]. Gender and age are insignificant. Political orientation, measured on an eleven-point scale ranging from 0 (left) to 10 (right) is highly significant with the expected sign: right-leaning subjects opt for less redistribution. Participants who performed better to the task are significantly less likely to ask for redistribution. The remaining models control for the influence of the few subjects who managed to become overachievers despite being assigned to the hard task or vice versa. In Specification (3), we include a dummy variable identifying the switchers, and in Specification (4) we estimate the models without the switchers. For both models, we find almost identical results as in Specification (2). Finally, in Specification (5) we instrument the overachiever status by the nature of the assigned task to estimate the marginal impacts accounting for the potential selection effect of the switchers.[10] Again we find that our results remain unchanged, giving us confidence that the observed effects are not driven by the switchers. These results are robust to separate regressions (see table A3) and to linear estimates[11].

4.3 Redistributive Taxation Game

Our second dependent variable measures the demand for redistribution affecting a subject's own future earnings. Figure 5 displays the average tax rates per treatment (for histograms of the tax rates see Figure A1 in the Appendix). First, we replicate the findings of Deffains et al. (2016) and find that, in the absence of information, underachievers ask for significantly more redistribution than overachievers do (42.7% vs. 37.6%; $p = 0.048$, ranksum test).

Result 3. *In absence of information, underachievers opt for higher redistributive tax rates than overachievers.*

---

[9]We compute scores of standardized performance by centering and standardizing the performance scores at the session and task level.

[10]To fully capture the treatment effect, we rely on an instrumental-variable approach such as suggested by Angrist et al. (1996). To do so, we instrument the actual treatment (overachiever or underachiever group) by the intention to treat (easy or hard task assignment). The instrumented variable *overachiever* is no longer correlated with the error term (due to the composition effect), because the task assignment was randomly allocated.

[11]Results available upon request.

Table 2: Regressions of the demand for redistribution in the RTG

Variable	Redistribution RTG				
	(1)	(2)	(3)	(4)	(5)
PostTask	0.0299	-0.0403	-0.0490	-0.0497	-0.0553
	(0.163)	(0.167)	(0.167)	(0.169)	(0.171)
PreTask	-0.414*	-0.531**	-0.539**	-0.502**	-0.479**
	(0.218)	(0.227)	(0.227)	(0.231)	(0.241)
Overachiever × *Baseline*	-0.309*	-0.306*	-0.299*	-0.283*	-0.310*
	(0.166)	(0.167)	(0.167)	(0.169)	(0.172)
Overachiever × *PostTask*	-0.437***	-0.420**	-0.409**	-0.414**	-0.395**
	(0.165)	(0.166)	(0.167)	(0.168)	(0.177)
Overachiever × *PreTask*	0.192	0.205	0.229	0.204	0.104
	(0.253)	(0.255)	(0.257)	(0.262)	(0.292)
Age		0.0225	0.0267	0.0341	0.022
		(0.0207)	(0.0209)	(0.0221)	(0.0207)
Political orientation		-0.0566**	-0.0605**	-0.0550**	-0.0572**
		(0.0249)	(0.0251)	(0.0256)	(0.0249)
Female		0.178*	0.185*	0.184*	0.179
		(0.108)	(0.109)	(0.110)	(0.108)
Standardized performance		-0.122**	-0.105*	-0.104*	-0.123**
		(0.0560)	(0.0586)	(0.0587)	(0.0583)
ME Overachiever in *Baseline*	-7.415*	-7.276*	-7.111*	-6.694*	-7.375*
	(3.966)	(3.960)	(3.973)	(3.993)	(4.091)
ME Overachiever in *PostTask*	-10.406**	-9.799**	-9.511**	-9.585**	-9.193**
	(3.898)	(3.841)	(3.852)	(3.863)	(4.104)
ME Overachiever in *PreTask*	4.435	4.590	5.128	4.569	2.326
	(5.842)	(5.697)	(5.746)	(5.860)	(6.530)
Switchers	.	.	Dummy	Excluded	IV
Observations	374	374	374	362	374
LL	-786.34	-778.78	-777.19	-752.68	-820.79

Standard errors in parentheses

Ordered probit estimates, and marginal effects (ME) in percentage points.

*** p<0.01, ** p<0.05, * p<0.1

Targets of the DDG excluded.

Second, similarly to the DDG we observe that the difference of redistribution choices between the two groups of players remains significant when information about the luck/effort ratio is provided *ex-post* ($p < 0.01$, ranksum test), but vanishes when the information comes *prior* to performing the real-effort task ($p = 0.521$, ranksum test). Regression analyses[12] show that the difference between overachievers and underachievers is significant at 1% in the *PostTask* treatment, and nonsignificant in the *PreTask* treatment (Columns 3 and 4 of Table 2). Again, we find that the information about the luck/effort ratio prevents the emergence of the self-serving bias only when the information is provided before the real-effort task. The difference between over- and underachievers is always negative, and (at least weakly) significant in *Baseline* and significant in *PostTask*. In *PreTask* the point estimate is positive and far from significance. These results hold for all our specifications controlling for the influence of the switchers. Last, the results are robust to separate regressions (see Table A4) and partially robust to linear specifications[13].

Result 4. *Overachievers and underachievers display similar redistribution choices in the RTG when information about the luck/effort ratio is provided prior to performing the real-effort task. Overachievers ask for significantly lower levels of redistribution than underachievers do when the information is provided after the real-effort task.*

5   Postdictions

Our hypotheses developed in Section 3.5 predicted a smaller or no bias in response to the additional information, but did not specify whether this should be due to changes in the behavior of overachievers or underachievers (or both). In the following, we present additional findings on the behavior of over- and underachievers. All findings that follow were not part of the hypotheses and are therefore postdictions.

We start by investigating how the bias was mitigated in *PreTask* relative to the other two treatments. As mentioned in the literature review, previous experimental works on redistribution concluded that individuals tend to reward effort and to compensate bad luck. In our setting, the clarification of the causal attribution is made by revealing to participants that luck was the predominant determinant of success and failure. This information should drive participants to ask for more redistribution as they would be more likely to think that the actual distribution of payoff results from luck.

The results observed for the DDG are in line with this theory. Figure 4 suggests that the reduction in the difference between overachievers and underachievers in the DDG is driven by the increase of the redistribution by overachievers. This intuition is supported by Wilcoxon rank-sum tests that detect no statistical difference in the redistribution levels chosen by underachievers across treatments ($p > 0.10$), but conclude that overachievers redistribute significantly more in the *PreTask* condition than in *Baseline* ($p = 0.004$), or in *PostTask* ($p = 0.012$). These results are also robust to multivariate analysis.

However, for the RTG we observe opposite effects. It appears from Figure 5 that overachievers tend to select similar redistribution levels across treatments, while underachievers opt for lower redistribution levels in *PreTask* relative to the other two treatments. Wilcoxon rank-sum tests only reject the null hypothesis when comparing underachievers under *PostTask* and *PreTask* ($p = 0.079$). Ordered probit estimations confirm this difference and show further that underachievers ask for significantly less taxation in *PreTask* than in *Baseline*.[14]

---

[12]All regressions exclude targets from the disinterested dictator game.

[13]The coefficient associated with overachievers for *Baseline* loses in significance. Results available upon request.

[14]All regression results are available on request.

6   Conclusion

We explored the role in information about "social mobility" on the political self-serving bias in redistributive preferences. On one hand we, found that high ambiguity in terms of the relative importance of external and internal factors is not a necessary condition for the political self-serving bias to arise. Participants exhibit such a bias (taking responsibility for success but blaming others for failure) even if we disclose that success and failure was almost exclusively due to chance, provided that the information comes *after* the realization of success or failure. Even with full transparency, the experience of success causes subjects to lower their willingness to redistribute, both in the disinterested dictator game and in the redistributive taxation game.

On the other hand, our experimental results suggest that the political self-serving bias does not arise when information about the importance of external versus internal factors is provided *before* the realization of failure or success. If subjects know the game is rigged before they engage in the real-effort task, then underachievers and overachievers display similar redistributive preferences in the disinterested dictator game, and the desired tax rates of underachievers and overachievers converge in the redistributive taxation game.

Inspired by the literature on association-based judgment errors (e.g., Arkes, 1991), we investigated whether objective information about social mobility can help bridging the gap between the two positions. Our results are somewhat pessimistic: The bias seems to be robust against ex-post information. This puts into question whether the dissemination of empirical results on the transmission of wealth and social mobility (e.g., Bowles and Gintis, 2002) can alter views on redistribution. On the other hand, our results suggest that providing the information ex-ante eradicates the systematic differences between under- and overachievers. The crucial question for a policy application is the meaning of *ex-ante*. A substantial determinant of one's position in a society depends on factors already present at birth. Consequently, if ex-ante is taken literally, an intervention like the one we apply in our treatment *PreTask* is not feasible. A less strict interpretation of ex-ante, on the other hand, lends itself to interventions in the educational system, where there is at least a partial veil of ignorance with respect to future success and failure.

While we focus on identifying the causal chain from unequal outcomes on redistributive preferences, it is likely that in reality the converse is also the case, namely that self-serving biases affect the distribution of outcomes, as discussed in Alesina et al. (2012). It would be interesting for further research to explore the dynamics of the self-serving bias in repeated interaction. Furthermore, while we have demonstrated that ex-post information does not seem to debias redistributive preferences, we cannot conclude that the bias is immune to any kind of information. For example, participants could be informed about the very existence of the political self-serving bias before making redistributive choices. Similarly, one could consider institutional settings where individuals must consider the viewpoint of the other members of their society: overachievers could change their views if they decide after contemplating the situation of the underachievers, and vice versa.

To conclude, our results replicate earlier findings on the political self-serving bias and explore two potentially debiasing interventions. We find that the bias is surprisingly robust against ex-post information. Our ex-ante information treatment, on the other hand, reveal no systematic differences between over- and underachievers. The observation that the bias survives our ex-post intervention is difficult to rationalize based on theories of social preferences or the accountability principle. This suggests that our setup produces biased behavior, best explained by assuming individuals are subject to belief perseverance. The second result is more positive for theories of rational choice and suggests that, in order to produce biased beliefs, we need some degree of ambiguity in the experimental setup. The mere experience of success and failure alone does not bias redistribution choices between under- and overachievers.

References

Ackert, L. F., Martinez-Vazquez, J., and Rider, M. (2007). Social preferences and tax policy design: Some experimental evidence. *Economic Inquiry*, 45(3):487–501.

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1):123–129.

Alesina, A. and Angeletos, G.-M. (2005). Fairness and Redistribution. *American Economic Review*, 95(4):960–980.

Alesina, A., Cozzi, G., and Mantovan, N. (2012). The evolution of ideology, fairness and redistribution. *Economic Journal*, 122(565):1244–1261.

Alesina, A. and Giuliano, P. (2011). Preferences for redistribution. In *Handbook of Social Economics*, volume 1A, chapter 4, pages 93–131. Elsevier B.V.

Anderson, C. A. (1982). Inoculation and Counterexplanation: Debiasing Techniques in the Perseverance of Social Theories. *Social Cognition*, 1(2):126–139.

Anderson, C. A. (1983). Abstract and Concrete Data in the Perseverance of Social Theories: When Weak Data Lead to Unshakeable Beliefs. *Journal of Experimental Social Psychology*, 19(2):93–108.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, 110(3):486–498.

Babcock, L. and Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives*, 11(1):109–126.

Babcock, L., Loewenstein, G., and Issacharoff, S. (1997). Creating convergence: Debiasing biased litigants. *Law & Social Inquiry*, 22(4):913–925.

Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. F. (1995). Biased judgments of fairness in bargaining. *American Economic Review*, 85(5):1337–1343.

Balafoutas, L., Kocher, M. G., Putterman, L., and Sutter, M. (2013). Equality, equity and incentives: An experiment. *European Economic Review*, 60:32–51.

Bellemare, C., Kröger, S., and van Soest, A. (2008). Measuring Inequity Aversion in a Heterogenous Population Using Exprimental Decisions and Subjective Probabilities. *Econometrica*, 76(4):815–839.

Bjerk, D. (2016). In front of and behind the veil of ignorance: An analysis of motivations for redistribution. *Social Choice and Welfare*, 47(4):791–824.

Bonica, A., McCarty, N., Poole, K. T., and Rosenthal, H. (2013). Why hasn't democracy slowed rising inequality? *Journal of Economic Perspectives*, 27(3):103–124.

Bowles, S. and Gintis, H. (2002). The inheritance of inequality. *Journal of Economic Perspectives*, 16(3):3–30.

Cabrales, A., Nagel, R., and Rodríguez Mora, J. V. (2012). It is Hobbes, not Rousseau: An experiment on voting and redistribution. *Experimental Economics*, 15(2):278–308.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.

Cappelen, A. W. and Tungodden, B. (2017). Fairness and the proportionality principle. *Social Choice and Welfare*, 49(3-4):709–719.

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

Croson, R. and Konow, J. (2009). Social preferences and moral biases. *Journal of Economic Behavior & Organization*, 69(3):201–212.

Deffains, B., Espinosa, R., and Thöni, C. (2016). Political self-serving bias and redistribution. *Journal of Public Economics*, 134:67–74.

Durante, R., Putterman, L., and van der Weele, J. (2014). Preferences for redistribution and perception of fairness: An experimental study. *Journal of the European Economic Association*, 12(4):1059–1086.

Eisenkopf, G., Fischbacher, U., and Föllmi-Heusi, F. (2013). Unequal opportunities and distributive justice. *Journal of Economic Behavior & Organization*, 93:51–61.

Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7):3330–3348.

Fehr, D. (2018). Is increasing inequality harmful? Experimental evidence. *Games and Economic Behavior*, 107:123–134.

Frohlich, N. and Oppenheimer, J. A. (1990). Choosing justice in experimental democracies with production. *American Political Science Review*, 84(2):461–477.

Gerber, A., Nicklisch, A., and Voigt, S. (2019). The role of ignorance in the emergence of redistribution. *Journal of Economic Behavior & Organization*, 163:239–261.

Gino, F., Norton, M. I., and Weber, R. A. (2016). Motivated bayesians: Feeling moral while acting egoistically. *Journal of Economic Perspectives*, 30(3):189–212.

Großer, J. and Reuben, E. (2013). Redistribution and market efficiency: An experimental study. *Journal of Public Economics*, 101:39–52.

Guenther, C. L. and Alicke, M. D. (2008). Self-Enhancement and Belief Perseverance. *Journal of Experimental Social Psychology*, 44:706–712.

Gächter, S. and Riedl, A. (2005). Moral property rights in bargaining with infeasible claims. *Management Science*, 51(2):249–263.

Hoffman, E. and Spitzer, M. L. (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *Journal of Legal Studies*, 14(2):259–297.

Jolls, C. and Sunstein, C. R. (2006). Debiasing through law. *Journal of Legal Studies*, 35(1):199–241.

Karadja, M., Mollerstrom, J., and Seim, D. (2017). Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics*, 99(2):201–212.

Karagözoğlu, E. and Riedl, A. (2015). Performance information, production uncertainty, and subjective entitlements in bargaining. *Management Science*, 61(11):2611–2626.

Kataria, M. and Montinari, N. (2012). Risk, entitlement and fairness bias: Explaining preferences for redistribution in multi-person setting. *Working Paper*.

Kittel, B., Kanitsar, G., and Traub, S. (2017). Knowledge, power, and self-interest. *Journal of Public Economics*, 150:39–52.

Klor, E. F. and Shayo, M. (2010). Social identity and preferences over redistribution. *Journal of Public Economics*, 94(3-4):269–278.

Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *The American Economic Review*, 90(4):1072–1091.

Konow, J. (2009). Is fairness in the eye of the beholder? An impartial spectator analysis of justice. *Social Choice and Welfare*, 33(1):101–127.

Konow, J., Saijo, T., and Akai, K. (2019). Equity versus equality: Spectators, stakeholders and groups. *Forthcoming in: Journal of Economic Psychology*.

Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980). Memory reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2):107–118.

Kovárík, J., Martínez-Macías, I., and Miller, L. (2018). Distributive preferences and effort provision: What determines what? *Working Paper*.

Larrick, R. P. (2004). Debiasing. *Blackwell handbook of judgment and decision making*, pages 316–338.

Lord, C. G., Lepper, M. R., and Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6):1231–1243.

Luhan, W. J., Poulsen, O., and Roos, M. W. M. (2019). Money or morality: Fairness ideals in unstructured bargaining. *Social Choice and Welfare*, 53(4):655–675.

Nisbett, R. E., Krantz, D. H., Jepson, C., and Kunda, Z. (1983). The use of statictical heuristics in everyday inductive reasoning. *Psychological Review*, 90(4):339–363.

Norton, E. C., Wang, H., and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal*, 4:154–167.

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

Ross, L., Lepper, M. D., and Hubbard, M. (1975). Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32(5):880–892.

Schildberg-Hörisch, H. (2010). Is the veil of ignorance only a concept about risk? An experiment. *Journal of Public Economics*, 94(11-12):1062–1066.

Slovic, P. and Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4):544–551.

Spiekermann, K. and Weiss, A. (2016). Objective and subjective compliance: A norm-based explanation of 'moral wiggle room'. *Games and Economic Behavior*, 96:170–183.

Tinghög, G., Andersson, D., and Västfjäll, D. (2017). Are individuals luck egalitarians? - an experiment on the influence of brute and option luck on social preferences. *Frontiers in Psychology*, 8:1–8.

Tyran, J.-R. and Sausgruber, R. (2006). A little fairness may induce a lot of redistribution in democracy. *European Economic Review*, 50(2):469–485.

Ubeda, P. (2014). The consistency of fairness rules: An experimental study. *Journal of Economic Psychology*, 41:88–100.

Weinzierl, M. (2017). Popular acceptance of inequality due to innate brute luck and support for classical benefit-based taxation. *Journal of Public Economics*, 155:54–63.

A   Appendix: Tables & Figures

Table A1: Descriptive Statistics per type of participant and per treatment.

	Baseline		PostTask		PreTask	
	Over	Under	Over	Under	Over	Under
Redistribution DDG	35.89	29.04	38.66	31.39	41.64	38.438
	(21)	(20.27)	(22.8)	(16.93)	(24.59)	(16.214)
Redistribution RTG	42.74	37.62	45.48	34.88	36.67	38.889
	(23.61)	(30.2)	(24.51)	(20.09)	(22.8)	(23.997)
age	19.71	20.06	20.96	20.87	22.19	22.167
	(1.33)	(1.85)	(3.22)	(2.33)	(3.5)	(3.094)
polit	4.9	4.46	4.35	4.12	4.11	3.278
	(2.25)	(2.2)	(2.22)	(2.07)	(1.6)	(2.25)
female	51%	48%	63%	55%	47%	50%

Means and standard deviations (in parentheses)

Figure A1: Histograms of the redistribution decisions in the RTG.
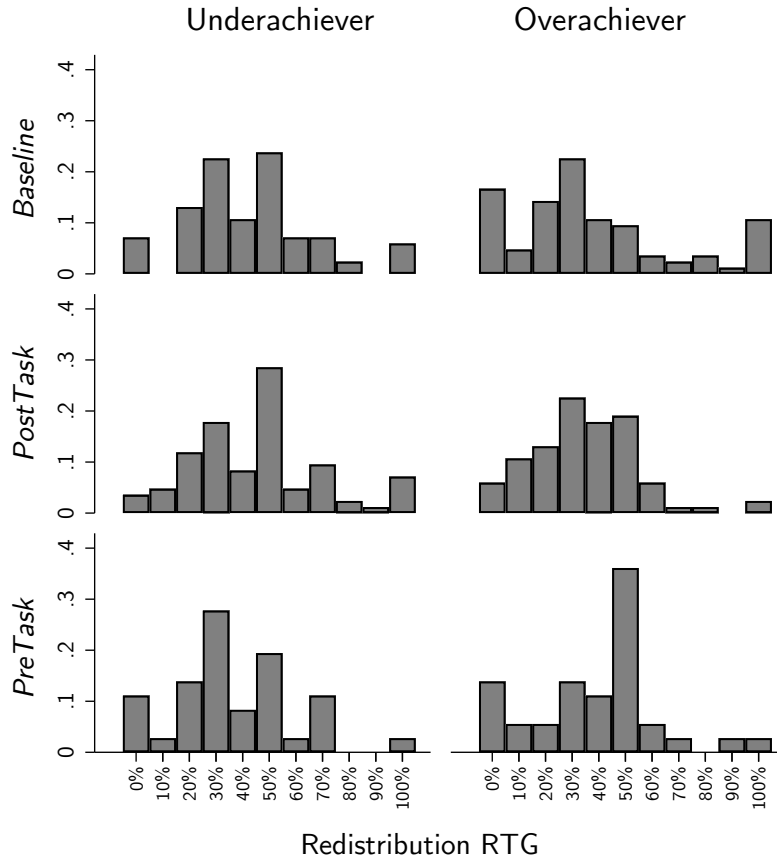
Table A2: Regressions of the absolute performance.

Variable	Performance	
	(1)	(2)
PostTask	-0.421	-0.441
	(0.348)	(0.354)
PreTask	0.662	0.510
	(0.449)	(0.463)
Overachiever $\times$ *Baseline*	6.202***	6.169***
	(0.348)	(0.348)
Overachiever $\times$ *PostTask*	6.226***	6.198***
	(0.348)	(0.348)
Overachiever $\times$ *PreTask*	6.278***	6.291***
	(0.531)	(0.532)
Age		0.0548
		(0.0450)
Political orientation		0.000483
		(0.0522)
Female		-0.407*
		(0.225)
Constant	10.35***	9.479***
	(0.246)	(0.975)
H_0: Over. $\times$ *Baseline*=Over. $\times$ *PostTask* (p-value)	0.9614	0.9538
H_0: Over. $\times$ *Baseline*=Over. $\times$ *PreTask* (p-value)	0.9055	0.8477
H_0: Over. $\times$ *PreTask*=Over. $\times$ *PostTask* (p-value)	0.9353	0.8831
Observations	408	408
R2	0.663	0.667

OLS, Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A3: Separate regressions of the demand for redistribution in the DDG.

Variable	Redistribution DDG				
	(1)	(2)	(3)	(4)	(5)
Regression for *Baseline*					
ME Overachiever	-6.723**	-6.426**	-6.540**	-6.587**	-7.742**
	(3.265)	(3.179)	(3.204)	(3.226)	(3.404)
Regression for *PostTask*					
ME Overachiever	-7.878**	-7.274**	-7.764**	-7.820**	-7.942**
	(3.157)	(3.004)	(3.107)	(3.097)	(3.246)
Regression for *PreTask*					
ME Overachiever	-0.732	-0.987	-1.766	-1.692	-3.071
	(4.621)	(4.232)	(4.649)	(4.853)	(5.219)
Controls	No	Yes	Yes	Yes	Yes
Switchers	.	.	Dummy	Excluded	IV

Standard errors in parentheses
Ordered probit estimates, and marginal effects (ME) in percentage points.
*** p<0.01, ** p<0.05, * p<0.1

Table A4: Separate regressions of the demand for redistribution in the RTG.

Variable	Redistribution RTG				
	(1)	(2)	(3)	(4)	(5)
Regression for *Baseline*					
ME Overachiever	-7.479*	-8.176*	-7.786*	-7.755*	-8.473*
	(4.332)	(4.239)	(4.285)	(4.258)	(4.387)
Regression for *PostTask*					
ME Overachiever	-10.520***	-9.576***	-9.225***	-9.132***	-8.909**
	(3.523)	(3.476)	(3.512)	(3.504)	(3.649)
Regression for *PreTask*					
ME Overachiever	4.306	4.643	4.714	4.597	3.869
	(5.740)	(5.803)	(5.855)	(5.927)	(6.603)
Controls	No	Yes	Yes	Yes	Yes
Switchers	.	.	Dummy	Excluded	IV

Standard errors in parentheses

Ordered probit estimates, and marginal effects (ME) in percentage points.

*** p<0.01, ** p<0.05, * p<0.1

Targets of the DDG excluded.

B   Marginal Effects

In Section 4, we are interested in estimating the marginal effect of being an overachiever in each treatment. To do so, we run ordered probit estimations where the equation for the latent utility writes:

$$y^* = \beta_1 posttask + \beta_2 pretask + \beta_3 over \times baseline + \beta_4 over \times pretask + \beta_5 over \times posttask + X\beta + \epsilon \tag{1}$$

where X is the matrix of control variables including a constant term.

In a linear model, the marginal effects of being an overachiever would be equal to β_3 for the *Baseline* treatment, β_4 for the *PreTask* treatment, and β_5 for the *PostTask* treatment. As Ai and Norton (2003); Norton et al. (2004) note, the marginal effect of an interaction of variables in non-linear models is more complex to retrieve. We derive below how we computed the marginal effects on redistribution for our ordered probit estimations.

Let y^* be the latent utility associated with the redistribution decision, where the individual can make a choice between J redistribution options, $y \in \{1, ..., J\}$. We have the following decision rule:

$$y = 1 \text{ if } y^* \leq c_1$$
$$y = 2 \text{ if } c_1 \leq y^* < c_2$$
$$...$$
$$y = j \text{ if } c_{j-1} \leq y^* < c_j$$
$$...$$
$$y = J - 1 \text{ if } c_{J-2} \leq y^* < c_{J-1}$$
$$y = J \text{ if } c_{J-1} \leq y^*$$

where y^* is defined as in equation 1. Alternatively, we write $y^* = Z\delta$ by stacking together the independent variables.

Assuming normally distributed random terms, the probability associated with each outcome is given by:

$$Pr[y = 1|Z] = \Phi(c_1 - Z\delta)$$
$$Pr[y = 2|Z] = \Phi(c_2 - Z\delta) - \Phi(c_1 - Z\delta)$$
$$...$$
$$Pr[y = j|Z] = \Phi(c_j - Z\delta) - \Phi(c_{j-1} - Z\delta) \tag{2}$$
$$...$$
$$Pr[y = J - 1|Z] = \Phi(c_{J-1} - Z\delta) - \Phi(c_{J-2} - Z\delta) \tag{3}$$
$$Pr[y = J|Z] = 1 - \Phi(c_{J-1} - Z\delta)$$

Baseline.   First, we derive the marginal effect of being an overachiever in the baseline treatment. The probability of making the choice j as an underachiever is:

$$\text{If } j = 1 : Pr[y = 1|Z, under, baseline] = \Phi(c_1 - X\beta)$$
$$\text{If } 1 < j < J : Pr[y = j|Z, under, baseline] = \Phi(c_j - X\beta) - \Phi(c_{j-1} - X\beta)$$
$$\text{If } j = J : Pr[y = J|Z, under, baseline] = 1 - \Phi(c_{j-1} - X\beta)$$

For an overachiever, the associated probabilities write:

$$\text{If } j = 1 : Pr[y = 1|Z, over, baseline] = \Phi(c_1 - X\beta - \beta_3)$$
$$\text{If } 1 < j < J : Pr[y = j|Z, over, baseline] = \Phi(c_j - X\beta - \beta_3) - \Phi(c_{j-1} - X\beta - \beta_3)$$
$$\text{If } j = J : Pr[y = J|Z, over, baseline] = 1 - \Phi(c_{j-1} - X\beta - \beta_3)$$

The change in probabilities for overachievers for each choice in the baseline treatment is given by $Pr[y = j|Z, over, baseline] - Pr[y = j|Z, under, baseline]$. For each specific outcome, it writes:

$$\text{If } j = 1 : ME_j(baseline) = \Phi(c_1 - X\beta - \beta_3) - \Phi(c_1 - X\beta)$$
$$\text{If } 1 < j < J : ME_j(baseline) = \Phi(c_j - X\beta - \beta_3) - \Phi(c_{j-1} - X\beta - \beta_3) - \Phi(c_j - X\beta) + \Phi(c_{j-1} - X\beta)$$
$$\text{If } j =< J : ME_j(baseline) = 1 - \Phi(c_{j-1} - X\beta - \beta_3) - 1 + \Phi(c_{j-1} - X\beta)$$

We thus obtain all changes in probability associated with each choice for the baseline treatment: $\forall j, ME_j(baseline)$.

The marginal effect of being an overachiever on the redistribution level is given by the sum of proposed redistribution outcomes times the increased probability of these redistribution outcomes, which writes:

$$ME(baseline) = \sum_{j=1}^{j=J} j \times ME_j(baseline) \tag{4}$$

PreTask. The computation of the marginal effects for the Pre-task treatment is similar up to a constant term β_2 which captures an overall treatment variation. We have:

$$\text{If } j = 1 : ME_j(pretask) = \Phi(c_1 - X\beta - \beta_2 - \beta_4) - \Phi(c_1 - X\beta - \beta_2)$$
$$\text{If } 1 < j < J : ME_j(pretask) = \Phi(c_j - X\beta - \beta_2 - \beta_4) - \Phi(c_{j-1} - X\beta - \beta_2 - \beta_4) - \Phi(c_j - X\beta - \beta_2) + \Phi(c_{j-1} - X\beta - \beta_2)$$
$$\text{If } j =< J : ME_j(pretask) = 1 - \Phi(c_{j-1} - X\beta - \beta_2 - \beta_4) - 1 + \Phi(c_{j-1} - X\beta - \beta_2)$$

PostTask. Similarly, the post-task marginal effects are given by:

$$\text{If } j = 1 : ME_j(posttask) = \Phi(c_1 - X\beta - \beta_1 - \beta_5) - \Phi(c_1 - X\beta - \beta_1)$$
$$\text{If } 1 < j < J : ME_j(posttask) = \Phi(c_j - X\beta - \beta_1 - \beta_5) - \Phi(c_{j-1} - X\beta - \beta_3 - \beta_5) - \Phi(c_j - X\beta - \beta_1) + \Phi(c_{j-1} - X\beta - \beta_1)$$
$$\text{If } j =< J : ME_j(posttask) = 1 - \Phi(c_{j-1} - X\beta - \beta_1 - \beta_5) - 1 + \Phi(c_{j-1} - X\beta - \beta_1)$$

C   Instructions

Instructions Part One

Hello everyone!

You are about to participate in an experiment, and we thank you for your participation. This experiment will be in two parts. These instructions are for the first part of the game. The instructions about the second part will be distributed at the end of the first part. If the instructions are not clear, or if a question persists, please raise your hand and wait for an instructor to come to you.

The present experiment consists of purely individual decision moments as well as group interactions with other individuals present in the room. At certain stages of the game, your decision may affect the gains of other participants, and vice versa. Thus, it is strictly forbidden to communicate for the duration of the experiment. In case of violation of this rule, we will be obliged to exclude you from the room.

Throughout this experiment, you will obtain ECU (*Experimental Currency Unit*). At the end of the experiment, the total number of ECUs you have accumulated throughout the game will be converted into EURO, and you will then be returned. The conversion rate is 5 ECU for 1 Euro.

The first part of this experiment is broken down into several stages. At first, you will be randomly assigned to a task. This task will consist in counting the number of 1s present in a sequence of 0s and 1s. This task will be timed, and your gains will depend on the number of correct answers you have provided (1 ECU per correct answer). The task that will be assigned to you can be either easy (50% chance) or difficult (50% chance). Each task type contains the same number of ECUs to win.

Example: 011010. This series of digits contains 3 ones, so the correct answer is: 3.

[The following paragraph is for PRETASK treatment only.] Note that at the end of this first task, the computer will tell you if you are among the 50% of the most successful participants, or the 50% of the least successful participants. In reality, **the experiment is made in such a way that it is <u>almost impossible</u> for a participant who has had the difficult task to do better than a participant who has had the easy task**. Indeed, all participants in previous sessions who had the hard task were consistently less successful than most participants in their session. In other words, none of the participants who had the difficult task was more successful than a participant who had the easy task.

Once all participants have completed the task assigned to them, you will need to answer a few questions. Then, two participants (target participants) will be drawn at random. For these two participants, the first part of the experiment will stop. Of these two selected participants, the participant with the most ECU will be called Participant A, while the participant with the least ECU will be called Participant B.

The other participants (judges) will not know the number of ECUs of these two selected players. They will, however, be able to redistribute the ECUs that player A has in addition to player B to player B. Each judge participant may choose not to redistribute, redistribute partially, or redistribute completely.

Once all the judging participants have made a redistribution choice, a solution will be drawn randomly and will be put in place for the two target candidates only.

Example: A judge participant decides to redistribute 18% of the surplus from player A to player B. When making a decision, the player judge does not know how much ECU player A has in addition to B. His decision is selected to be implemented. Player A actually has 11 ECUs more than Player B. Thus, 2 ECUs are taken from Player A to be given to Player B.

To summarize, the first stage of the game is as follows:
1) All participants receive a random assignment;
2) The participants carry out their task;
3) Participants answer a few questions;
4) Two participants are drawn at random (target participants);
5) Judge participants decide how to redistribute the ECUs that Player A has in addition to Player B;
6) One of the redistribution proposals is drawn and put in place: the judge players keep their money from the task, the two target players get the money corrected for the redistribution.

Instructions Part Two

The second part of the experiment is about to begin. In this part, you will have to complete several tasks. These tasks are substantially different from those presented to you in the first part of this experiment.

At the beginning of the game, you will be put in a group of 4 participants. The tasks to be performed will be individual, but all members present at the experiment will face the same tasks. From these tasks, you will get an individual income. This individual income may be taxed in part to be redistributed equally to all other members of the group (including you). All individuals in the group will be taxed with the same percentage. The tax rate can be: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%.

Example: The table below shows the income of fictitious participants and the effects of a tax rate of 60%. Participant A had 10 ECUs per task. The tax rate of 60% therefore removes 6 ECU. The total amount obtained by taxes is 36 ECU. Each participant receives an equal share of the 36 ECUs; participant A thus receives 9 ECUs. His final income is equal to his individual income (10 ECU) less what he costs (6 ECU) plus redistribution (9 ECU), for a total of 13 ECU.

	Individual income for the task	Level of taxation	Gains from redistribution	Final payoff
Participant A	*10*	*10*0.6=6*	*36 ÷ 4 = 9*	*10-6+9=13*
Participant B	*25*	*25*0.6=15*	*36 ÷ 4 = 9*	*25-15+9=19*
Participant C	*20*	*20*0.6=12*	*36 ÷ 4 = 9*	*20-12+9=17*
Participant D	*5*	*5*0.6=3*	*36 ÷ 4 = 9*	*5-3+9=11*
Sum	*60*	*36*	*36*	*60*

The redistribution rule can be determined in two different ways. On the one hand, you will be asked to indicate your preferred level of redistribution. Your answer will be strictly private and will not be communicated to other participants. On the other hand, you will have to choose a level of redistribution unanimously with the other members of your group.

When making a collective decision, you have five rounds to unanimously agree on a decision rule with the other three players in your group. Each turn, you will indicate the amount of taxation you wish to have. In the next round, you will learn the choice of each of the other participants in your group. The decision is made if all participants have indicated the same tax rate. In the event of a failure to make an unanimous decision, the participants will not play the second game and will receive a lump sum, which will be indicated to you at the beginning of the collective decision-making.

If your group has managed to agree on a common decision rule, you will play the second game. In this game you can win up to 80 ECU. The redistribution rule chosen will have a 50% chance of being the rule decided unanimously by the group. In the remaining 50% of cases, one of the private solutions will be randomly drawn to become the group's decision rule.

D   Appendix: Screenshots

Figure D1: Screenshot of the Control Questions for the Redistributive Taxation Game [1/4].

Figure D2: Screenshot of the Control Questions for the Redistributive Taxation Game [2/4].

Figure D3: Screenshot of the Control Questions for the Redistributive Taxation Game [3/4].

Avant de voter sur les règles de redistribution, un petit exemple pour s'assurer que les règles de redistribution sont claires.

Complétez le tableau suivant en fonction de la règle de distribution. (On imagine que les chocs peuvent modifier les profits de 5 ECU.

	Revenu initial (avant choc)	Revenu (après choc)	Montant prélevé avec taxe de 20%	Montant reçu par redistribution	Revenu final
A	5	5	1	2	6
B	5	0	0	2	
C	20	15	3	2	
D	20	20	4	2	
Somme	50	40	8	8	

OK

Figure D4: Screenshot of the Control Questions for the Redistributive Taxation Game [4/4].

Avant de voter sur les règles de redistribution, un petit exemple pour s'assurer que les règles de redistribution sont claires.

Complétez le tableau suivant en fonction de la règle de distribution. (On imagine que les chocs peuvent modifier les profits de 5 ECU.

	Revenu initial (avant choc)	Revenu (après choc)	Montant prélevé avec taxe de 20%	Montant reçu par redistribution	Revenu final
A	5	5	1	2	6
B	5	0	0	2	2
C	20	15	3	2	14
D	20	20	4	2	18
Somme	50	40	8	8	40

Fin de l'exemple

Figure D5: Screenshot of the Private Choice in the Redistributive Taxation Game.

Figure D6: Screenshot of the First Round of Bargaining in the Redistributive Taxation Game.

Figure D7: Screenshot of the Second Round of Bargaining in the Redistributive Taxation Game.