EXCESSIVE PAY IS NOT ABOUT THE NUMBERS
HOW POWER ABUSE ERODES INEQUALITY ACCEPTANCE*

Tora Knutsen[1] and Sonja Kovacevic[2]

[1]Institute for Social Research, Oslo and Nordic Institute for Studies in Innovation, Research and Education, Oslo. t.k.knutsen@samfunnsforskning.no
[2]University of Cologne. kovacevic@wiso.uni-koeln.de

Abstract

We study whether people judge the fairness of decisions by the outcome itself or by who makes them. In a large, pre-registered online experiment, we let impartial spectators decide final earnings after either a self-interested worker or a third party has chosen initial earnings. We match worker pairs to third-party decisions so that spectators see identical initial decisions and performances across treatments. Spectators assign 0.22 sd lower earnings to workers who determine their own initial earnings, compared to workers with identical but externally determined initial earnings. Cases where workers allocate themselves earnings exceeding their relative productivity fully account for this earnings reduction, suggesting that perceived power abuse, and not the mere presence of decision-making power, erodes inequality acceptance. Spectators show remarkable leniency for a broad range of worker choices, indicating that spectators reserve redistribution for power abuse where workers' choices defy all established fairness norms.

JEL classification: C91; D63; H23.

*We would like to thank Gøril Andreassen, Kai Barron, Kjell Arne Brekke, Alexander Cappelen, Dirk Engelmann, Tilman Fries, Lea Heursen, Steffen Huck, Bernd Irlenbusch, Andreas Kotsadam, Jo Thori Lind, Thomas McKay, Kalle Moene, Karine Nyborg, Henning Øien, Pia Pinger, Sonja Settele, Siv-Elisabeth Skjelbred, Erik Sørensen, Robert Stueber, Matthias Sutter, Bertil Tungodden, Vegard Wiborg, Johannes Wohlfahrt and various seminar participants for helpful comments and suggestions. This project received funding from "Småforsk" at the Department of Economics at the University of Oslo, Institute for Social Research through The Norwegian research council project 10152 and the Skaug fund at Statistics Norway. We also acknowledge support from UiO:Nordic.

1 Introduction

Top CEOs are overpaid, while top musicians deserve their equally high pay - observers insist. Star performers such as Taylor Swift encounter little criticism for their concert tour revenues, while the compensation package of top executives such as Sundar Pichai provokes negative headlines. Perceived decision-making power may affect what observers find fair. Executive pay involves board negotiations where asymmetric information may allow CEOs some discretion. Star musicians' earnings, relying on consumption choices, leave less apparent room for individual discretion.

Research on inequality acceptance has increasingly focused on the earnings process, rather than inequality per se, addressing a plethora of allocation rules. However, this literature has largely overlooked the role of decision-making power wielded by the self-interested parties. In real-world settings, such discretion may shape final earnings. This paper examines how spectators evaluate fairness when stakeholders hold decision-making power, and whether spectators object to discretionary power in general, or only when decision-makers exploit their position by making choices that violate established fairness norms.

We conduct a large-scale pre-registered experiment with 6000 participants on Prolific to investigate how decision-making power affects the perceived fairness of inequality. We measure fairness perceptions through redistribution choices made by participants acting as impartial spectators. In observational settings, decision-making power may lead to higher inequality, complicating the distinction between inequality aversion and concerns about power abuse. In addition, verifying whether discretion caused a particular allocation remains challenging. For example, did the CEO earning high bonuses successfully steer their company through a crisis or did they merely persuade the board to adopt a favorable performance measure? Our experimental setting allows us to disentangle objections to power abuse from concerns about the inequality level.

Our experiment consists of two stages (Almås, Cappelen, and Tungodden 2020), a worker stage and a spectator stage. In the worker stage, we pair participants and ask them to complete an effort task. After completing the task, we randomly select one worker in each pair to make a decision about how to distribute joint earnings.

In the spectator stage, we present an impartial spectator with the relative performances and initial distribution from a worker pair. The spectator then makes a final distribution decision. We randomly assign spectators to one of two treatments that vary in who made the initial decision. In the Dictator Decides treatment, spectators learn that a worker in the pair made the decision. In the External Decision treatment, spectators learn that a third party made the decision.

Thus, across treatment conditions we vary whether a self-interested worker or a third party made the initial decision. We match worker pairs by performance so that spectators in each condition see worker pairs with identical performances. We also ensure that spectators see identical initial decisions across treatment conditions by using the initial decision of a deciding worker in a worker pair not only for their own pair, but also for the performance-matched worker pair. To spectators, we refer to this transferred decision as a third-party decision. The difference between spectators' choices when facing a deciding worker versus a third party identifies the causal effect of decision-making power on spectators' fairness perceptions.

This setting allows us to study how observers evaluate potential conflicts of interest in economic allocations. Specifically, we distinguish between two ways spectators might react negatively to stakeholder decision-making power: a process-oriented spectator may be concerned about the inherent potential for abuse, while a spectator with a perceived-intent concern may focus on whether the power was actually used in a self-serving manner. When deciding about a given allocation, spectators may either use their own fairness ideal as benchmark, that is, how they themselves would ideally split earnings, or only redistribute

when decisions fall outside the range of generally justifiable decisions.

We find that decision-making power matters: spectators assign on average 4 percentage points less to stakeholders (DD treatment) deciding their own initial earnings, compared to identical allocations set by an external party (ED treatment). This average effect masks strong heterogeneity, with negative reactions highly conditional on the type of allocation chosen by the dictator, pointing towards intent-based, rather than process-based, fairness views. Spectators redistribute heavily when they face stakeholders who take an earnings share exceeding their performance, assigning on average 6.5 percentage points lower earnings compared to the same choice made by a third party. In contrast, spectators show tolerance for allocations aligned with common fairness norms like meritocracy or egalitarianism, treating them similarly regardless of whether they were chosen by a stakeholder or a third party. This highlights that concerns about (un)fairness are primarily triggered by the perceived abuse of decision-making power, rather than its mere presence.

When spectators face choices outside established fairness norms, a significant fraction of spectators impose punitive distributions, assigning final earnings below the dictator's performance share (on average, this punishment reduces the dictator's share by 19.5 percentage points below performance). This contrasts sharply with the ED treatment, where similar overproportional allocations are typically corrected back towards the performance share. We find that spectators switching from correcting, towards punishing an initial decision by assigning shares well below the performance shares, tend to self-classify as politically left-leaning and have high socio-economic status.

This paper contributes to the large literature examining how the perceived source of inequality shapes fairness views and redistribution preferences. People's fairness views define the political culture of a society and determine which welfare systems can be sustained (Kuziemko et al. 2015; Alesina, Stantcheva, and Teso 2018; Alesina and La Ferrara 2005; Hvidberg, Kreiner, and Stantcheva 2023; Stantcheva 2021; Fehr, Epper, and Senn 2024).

Fairness assessments do not simply depend on the distribution of income or wealth in a society, but on the process leading to that distribution; people tend to accept inequalities resulting from effort or merit more than inequalities resulting from luck (e.g., Alesina and La Ferrara 2005; Cappelen et al. 2007; Almås, Cappelen, and Tungodden 2020; Cappelen et al. 2023). In addition, fairness views depend on choices about risk and social preferences, whether related (Akbaş, Ariely, and Yuksel 2019) or unrelated (Gärtner, Mollerstrom, and Seim 2017; Mollerstrom, Reme, and Sørensen 2015) to the choice at hand. Unlike merit and luck, decision-making power in the income process held by a party with a direct conflict of interest has received less attention.

Our primary contribution is to experimentally isolate the causal effect of decision-making on inequality acceptance. Unlike prior work showing that suspected cheating reduces inequality acceptance (Klimm 2019; Bortolotti et al. 2023), our experiment explicitly permits the use of power. By authorizing power use, our setting links to real-world allocation-processes where decision-makers wield power legally, even if the (potentially self-serving) use of that power and the resulting inequality may not be perceived as legitimate.

Decision-making power matters for allocations in both economic and political spheres. Our work offers a new perspective on the recent renewed interest in wage-setting power in labor markets (for example Card 2022; Berger, Herkenhoff, and Mongey 2022; Azar, Marinescu, and Steinbaum 2022). This research shows that firms may possess significant wage-setting power, suggesting that observed wages may reflect not only productivity but also the exercise of this power. Our experimental results show that the abuse of decision-making power significantly reduces inequality acceptance. This implies that outside observers may perceive wage disparities as less fair when they result from employers' wage-setting power rather than solely from productivity differences.

Second, we advance the understanding of how individuals evaluate fairness in situations involving potential conflicts of interest. Spectators do not exhibit a purely process-based con-

cern; they do not uniformly penalize stakeholder control over allocations. Instead, their redistribution decisions depend on the specific allocation chosen by the stakeholder. Crucially, these evaluations appear grounded in a tolerance for outcomes that align with a plurality of widely recognized fairness norms, primarily meritocracy and egalitarianism (consistent with findings on fairness pluralism, e.g., Cappelen et al. 2007). Spectators intervene only when allocations violate this set of acceptable norms. This suggests that they apply a 'normative range' rather than a single benchmark.

Finally, our findings suggest that spectators' normative range includes both equal splits (egalitarian) and proportional (meritocratic) pay, but excludes tournament pay (cases where higher performers take more than their performance share) and any allocation in which the lower performer claims a majority share. Such cases result in some spectators "punishing" the decider by allocating them a final wage below their performance share. While related literature examines third-party intervention and punishment across various domains (e.g., Fehr and Gächter 2002; Fehr, Fischbacher, and Gächter 2002; Ackfeld and Ockenfels 2021), we demonstrate its specific application to allocation decisions in which the decider has a direct stake.

In the next section, we present the conceptual discussion that frames our study's focus process-based, as opposed to intent-based, fairness judgments. Following this, we detail the experimental design and provide an overview of the data. In the results section, we first describe general features of spectator behavior and then present the main specification. We then explore the mechanism and investigate heterogeneity in spectator choices between the two treatments, before concluding in the final section.

2 Conceptual Discussion

Consider the example of a middle manager deciding on bonus payments for their team and themselves based on a performance criterion that they themselves define. This article

explores how spectators judge fairness in a setting with decision-making power, where we measure fairness perceptions through spectators´ redistribution decisions. We investigate whether spectators hold decision-makers accountable, and if so, whether they redistribute whenever decision-makers could potentially act self-servingly, or only when they actually make self-serving choices.

In our experiment, two workers contribute to a joint output with differing performance levels, and we randomly select one of the workers to decide how to allocate bonuses for both themselves and their coworker from a fixed bonus pool. This setup inherently creates a conflict of interest, as the decision-maker can choose to allocate a larger share to themselves than what is justified under any fairness norm.

Spectators deciding whether to redistribute could judge fairness at the level of either process or choice. A choice may align with a specific fairness principle (e.g., egalitarian, meritocratic, winner-takes-all) but the spectator may perceive the allocation process, where a party with a direct stake holds decision-making power, as inherently problematic due to the potential for self-serving bias. We test whether spectators prioritize process, redistributing more when decision-making power is present irrespective of the actual decisions, or whether they prioritize the decision-maker's perceived intent, redistributing on a case-by-case basis.

Process-oriented spectators focus on the integrity of the allocation procedure. They object to the fact that a stakeholder holds decision-making power, perceiving the inherent conflict of interest as problematic. For these spectators, the mere possibility of power abuse taints the outcome. They might redistribute simply because the allocation was determined by a self-interested party, even if the resulting allocation aligns with a recognized fairness norm. Their concern is rooted in the potential for power abuse, irrespective of whether it demonstrably occurred in the specific instance. This view suggests a higher propensity to intervene and redistribute in settings where decision power rests with a stakeholder compared to settings where it rests with a third party.

Perceived-intent spectators, by contrast, base their judgment on whether they consider the chosen allocation as an abuse of power. This view rules out blanket redistribution when stakeholders assign initial wages. When evaluating perceived intent, spectators may either use their own fairness view as the benchmark, or adopt a broader approach checking whether this decision could have resulted from a stakeholder staying true to *their* fairness norm.

We used a pilot study to elicit spectators' own fairness ideal points, finding that 3 out of 4 spectators held a meritocratic ideal point.[1] We pre-registered to test the coefficient on the interaction between the treatment and dummies for taking more than the meritocratic share to test whether spectators implement the predominantly meritocratic ideal point.

If spectators employ a *narrow* definition of power abuse, considering only choices outside all generally accepted fairness norms as power abuse, we would expect spectators to tolerate egalitarian, meritocratic and potentially also tournament pay allocations. Spectators would then reserve redistribution for choices that do not align with any accepted fairness norms, such as the lower performer choosing a higher share. If spectators instead employ a *broad* definition of power abuse, they may object to all allocations that deviate from their own fairness view. We investigate whether spectators follow a broad or narrow definition in practice by distinguishing between different worker choice types; egalitarian, meritocratic, tournament pay and evident power abuse, to see which allocations spectators accept.

3 Experimental Design

We conducted a two-stage experiment with two types of participants: workers and spectators. In the worker stage, one set of participants ("workers") did a task, and in the

1. In the pilot, we elicited spectators' choices for final earnings given worker performances without a worker-level decision stage. We did not elicit spectators' ideal point in the main experiment to avoid anchoring spectators' fairness decisions too heavily. In this pilot study with the same Prolific population, we find that 74% allocate payments equal to the performance share and 14% split the payment equally between the workers.

spectator stage, another set of participants ("spectators") made distribution choices.[2] This two-stage set-up, with participants in the role of impartial spectators, making real decisions on earnings for another group of participants, is a workhorse set-up in the experimental literature on redistribution preferences (for example Cappelen et al. 2023; Almås, Cappelen, and Tungodden 2020; Bortolotti et al. 2023; Klimm 2019). The study was pre-registered in the AEA RCT-registry.[3]

We ran both stages of the experiment online and recruited participants via Prolific.[4] Participants learned that their answers would be collected anonymously, that they could opt out of the study at any point and that we would not use deception.

3.1 Worker stage

We recruited 4000 participants for the worker stage, Because some participants dropping out and some worker pairs not being matched, we ended up with a final sample of 3924 workers in 1962 worker pairs.

Workers received a 1 GBP participation fee and were told that they could earn additional money in the experiment. On entering the experiment, workers learned that they were matched with another participant and had to complete a real-effort task for five minutes. The task consisted of counting zeroes in progressively longer number series, where the goal was to count correctly in as many series as possible (see Figure 1 for an example).

We showed workers their own and the other workers' relative performances after they

2. We programmed the experiment in Lioness, a web-based platform for interactive online experiments (Arechar, Gächter, and Molleman 2018).

3. We pre-specified the worker stage (Knutsen and Kovacevic 2021b) and spectator stage (Knutsen and Kovacevic 2021a) seperately.

4. The platform differs in some important ways from the more commonly used Amazon MTurk: The platform requires researchers to pay the minimum wage and as it pays participants through Paypal, the subject pool is more diverse in terms of nationality than MTurk which is restricted to certain countries. Prolific also requires the researcher to report to Prolific whenever they want to reject a submission. According to Palan and Schitter (2018), this can help avoiding a researcher-demand effect where participants might want to give the researcher the answer they think the researcher wants in order not to get rejected. The benefit of recruiting participants on an online platform is both flexibility and lower price at little cost to quality of responses (Litman, Robinson, and Rosenzweig 2015)

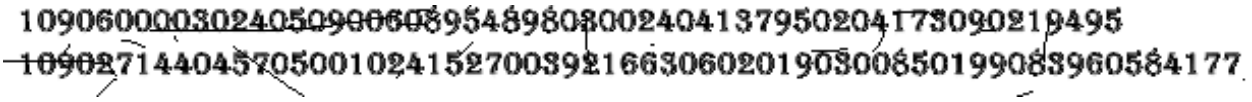


Figure 1: Example of number series shown to workers.

completed the task. Then, one worker (the "dictator") was randomly selected to allocate 2 GBP between the two. The dictator learned that their choice would be implemented with a 50% chance and that under no circumstances would their submission be rejected or would their coworker be able to review their choices.[5] However, we did not specify to dictators what would happen if their choice was not selected for implementation.[6] Workers received payments within a few days after spectators had made decisions for the 50% of randomly selected cases where their choice was implemented.

3.2 Performance Matching

Spectators across the two treatments see worker pairs identical in performance and initial wage, only differing in who determined the initial wage.

In order to obtain two identical groups with identical performances and initial choices, we divide worker pairs into two groups by matching performance distributions. Collecting all worker pairs where two or a multiple of two worker pairs had the same performance, we arrange worker pairs into two groups with an identical relative performance share distribution. Each worker pair in the first group thus has a corresponding "twin" in the second group.[7]

We equalize initial earnings by using choices made in the first group to set initial wages

5. Researchers can reject participants' submissions for completing surveys exceptionally fast or for answering gibberish. However, some participants are concerned about potential rejections since this affects their rating. We wanted to make sure that participants could freely chose their desired pay-out distribution.

6. If they anticipated that there would be a third party reviewing their choice, workers might have behaved more pro-social than they would have with a 100% implementation chance. Whether or not this is the case does not matter for this study as we are interested in spectators' reactions to workers' behavior and not workers' choices by themselves.

7. A total of 14 worker pairs had a unique value and could not be matched. We removed these worker pairs from the remainder of the experiment.

for performance-corresponding worker pairs in the second group. Thus, the dictator decides initial earnings for two worker pairs: first, in their own pair and second, for a worker pair in the control group with the same relative performance shares (Figure 2). We refer to the commuted dictator decision for the twin worker pair in the other group as the decision of a third party. The resulting distribution of performances and initial earnings is identical across the two groups. This allows us to isolate the treatment effect of the power position of dictators when analyzing spectator redistribution choices across groups.

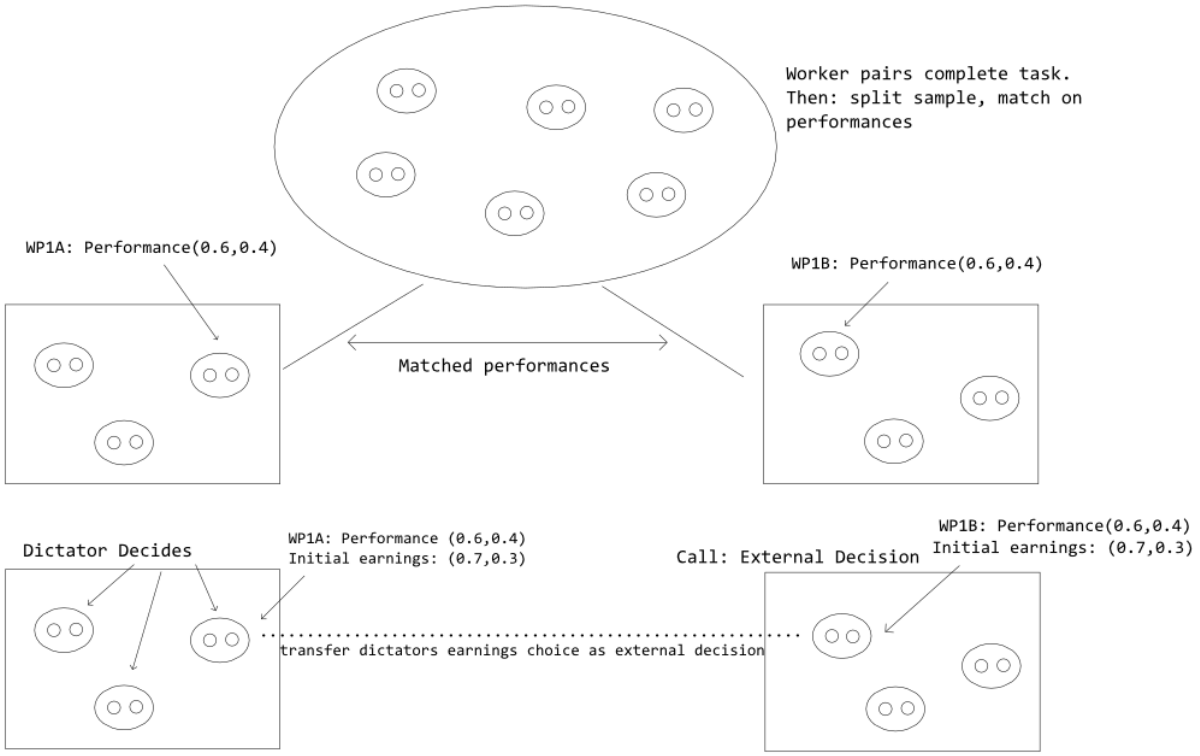


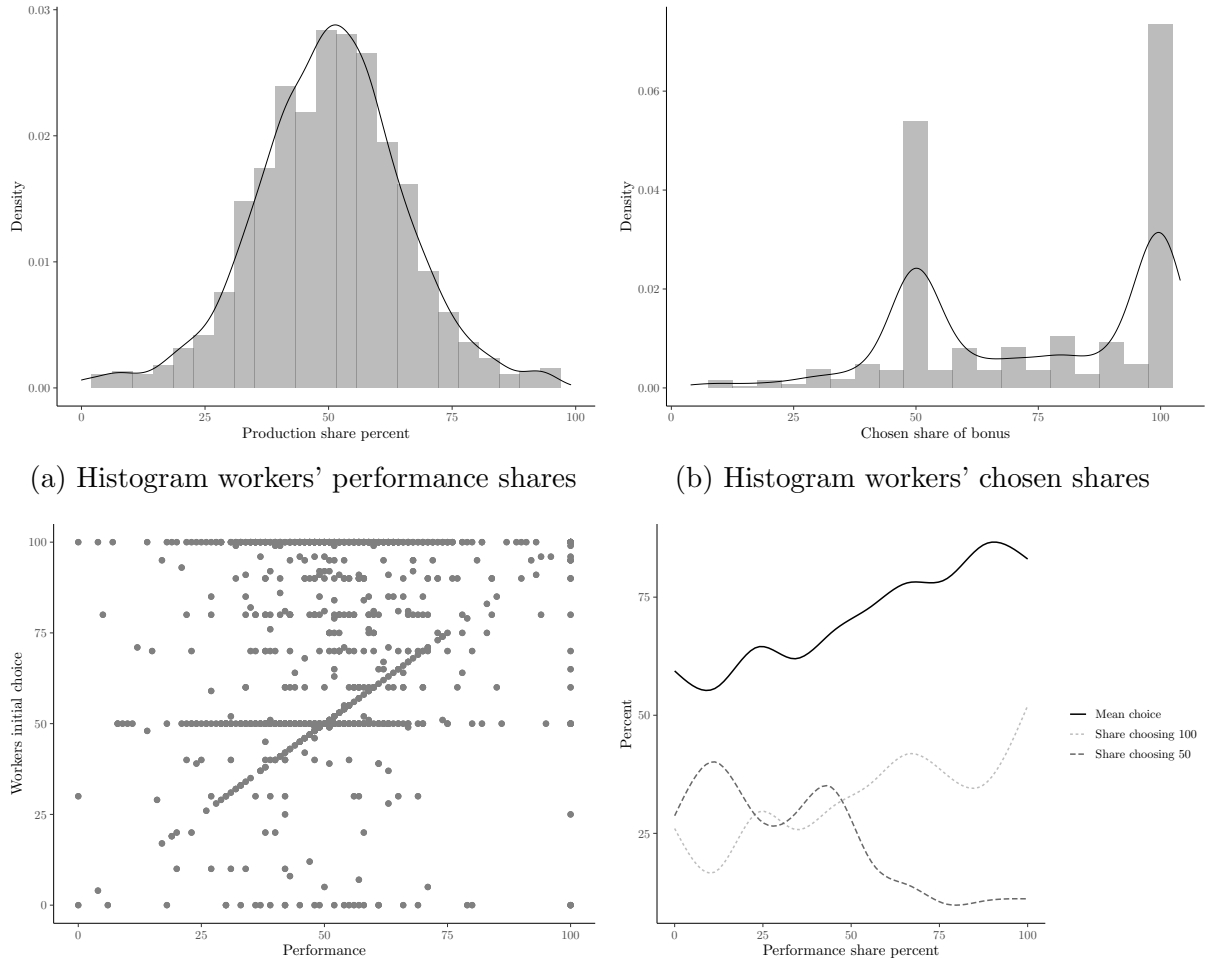
Figure 2: Constructing performance-inequality matched worker samples.

Figure 3 shows that worker performance shares follow a normal distribution.[8] The average dictator chooses an earnings share 18 percentage points above their performance share, but there is considerable heterogeneity in workers' choices. Taking all earnings and splitting them equally are the two most common choices. Figure 3 shows that claiming earnings equal

8. We observe a slightly elevated fraction of 0/100% pairs. These are cases where one participant did not spend time on the tasks.

to one's performance is a third common pattern of behaviour. The average choice lies well above 50% across all performance levels. While the propensity to claim 100% increases only moderately with performance, the propensity to split equally decreases drastically when performance exceeds 50%. In total 66% of dictators claim earnings exceeding their performance share.

Figure 3: Workers' performance shares and chosen shares.



(a) Histogram workers' performance shares

(b) Histogram workers' chosen shares

(c) Scatter plot workers' chosen shares and performance

(d) Mean chosen share and share of workers choosing 50 and 100

Notes: All figures show raw data from the first stage of the experiment. Panel (a) shows all workers' shares of production. Panel (b) shows the choices for the workers could choose their own earnings share. Panel (c) is a scatter plot of performance and choices. Panel (d) shows how the average choice and shares choosing 50%/100% vary with performance.

3.3 Spectator stage

We recruited 1962 spectators, one for each worker pair, between October 21 and 25, 2021. We invited participants from several different regions of the world and opened invitation slots to reach our pre-specified targets for each region.[9] Within each region, we excluded participants who had taken part in one of our two pilots.

On entering the experiment, spectators learn which task workers completed, their performances and the initial wages of the workers (distributive situation) and that they will make decisions for two worker pairs. Then, spectators are randomized into one of two treatments. In the Dictator Decides (DD) treatment, spectators learn that a worker *within* the pair had decided their own earnings share. In the External Decision (ED) treatment, spectators learn that a *third party* had decided earnings shares for both workers. We do not tell spectators anything about the identity of this third party.

As is common in online experiments, we conducted comprehension checks to ensure that subjects were attentive.[10] The instructions emphasized that spectators were making choices for real people, adding information on how and when workers would be paid. Spectators then chose final earnings in two worker pairs under the same treatment condition. Moving a slider allowed spectators to allocate relative earnings between the two workers. Spectators then proceeded to rationalize their choices in an open text question. Finally, spectators completed a set of background questions regarding political and redistribution preferences as well as their own economic status.[11]

Spectators were balanced across the treatment groups in terms of age, gender, employment status, nationality and student status as shown in Table 1. By design, the matched variables, performance and initial choice, are identically distributed across the two groups.

9. See Appendix B.3 for a detailed description of the regions included and behavior differences by country.

10. Subjects could not continue to the next page until correctly answering the comprehension questions, but could look up information about the experiment set-up as shown in Appendix A.2.

11. Appendix A.1 includes the full set of screenshots from the experiment. The background questions are listed in Appendix A.3

Variables collected after spectators received treatment may differ across conditions. Spectators in the DD treatment use more time on the study than ED treatment spectators and this difference is significant at the 1% level. Notably, spectators spent most of their extra time on the two redistribution decisions, indicating that spectators may perceive distributional fairness as more complex when dictators are able use their decision-making power for their own benefit. Variables on political views and economic status (where one places oneself economically on a ladder) are balanced across the two groups, indicating that our treatment did not impact answers other than to the redistribution choices.

Table 1: Summary statistics: Balance between DD and ED treatment groups

	Dictator		Assigned		
Variable	Mean	sd	Mean	sd	p-value
Matched variables					
Performance	52.77	18.31	52.77	18.31	1.00
Initial choice	71.52	27.56	71.52	27.56	1.00
Pre-choice variables					
Female	0.58	0.49	0.58	0.49	0.90
Batch Prolific	5.36	2.53	5.35	2.51	0.90
Age	27.87	8.92	28.11	9.27	0.57
Employed	0.57	0.50	0.55	0.50	0.37
Student	0.48	0.50	0.49	0.50	0.68
Post-choice variables					
Placement on left-right scale (1-4)	2.05	0.90	2.06	0.87	0.98
Placement on socio-economic ladder (1-10)	5.49	1.60	5.50	1.65	0.86
Time taken	654.02	322.51	615.92	287.77	0.01
N	981		981		

3.4 Treatments

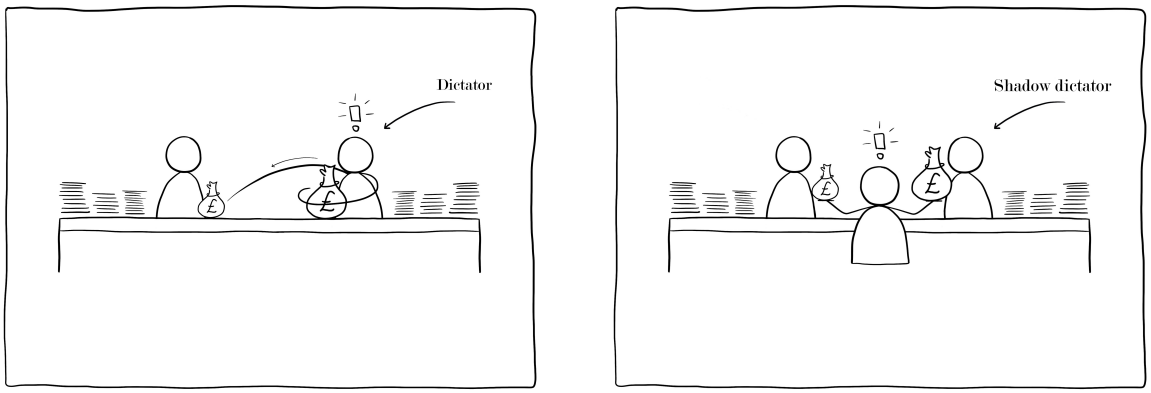
By construction, spectators face identical distributions of performances and initial earnings in each treatment condition. However, who decides initial earnings differs across treatments. In the DD treatment, spectators learn that a worker in the pair could decide earnings shares for both workers. In the ED treatment, spectators learn that a third party could decide earnings shares for both workers. In the following, we refer to the worker matched with

a performance-corresponding dictator as the dictator counterpart. Spectators receive this information through instructions and illustrations (Figure 4).

Instructions

- Dictator Decides (DD) treatment: *We let one of the participants working on the questions choose preliminary pay-offs. That is, they get to choose both how much to give to themselves and how much to give to the other participant. In this pair, participant A got to decide and made the following choice: "Give 80% of the total pay-off to myself (participant A, right side) and give 20% of the pay-off to participant B (left side).*
- External Decision (ED) treatment: *We let another participant choose preliminary pay-offs. That is, they get to choose both how much to give to participant A and how much to give to participant B. This third party made the following choice: "Give 80% of the total pay-off to participant A (right side) and give 20% of the pay-off to participant B (left side)."*

Illustration



(a) Dictator Decides (DD) treatment (b) External Decision (ED) treatment

Figure 4: Illustrations shown to spectators. Note that the sizes of the money bags are proportional to the split.

By comparing spectators' redistribution choices in the DD treatment and the ED decides treatment, we can identify the causal effect of decision-making power in the income process on spectators' fairness considerations. Moreover, we can identify whether spectators react to power in the income process per se, or whether they primarily react when deciding workers abuse their power to achieve earnings above their performance share.

4 Results

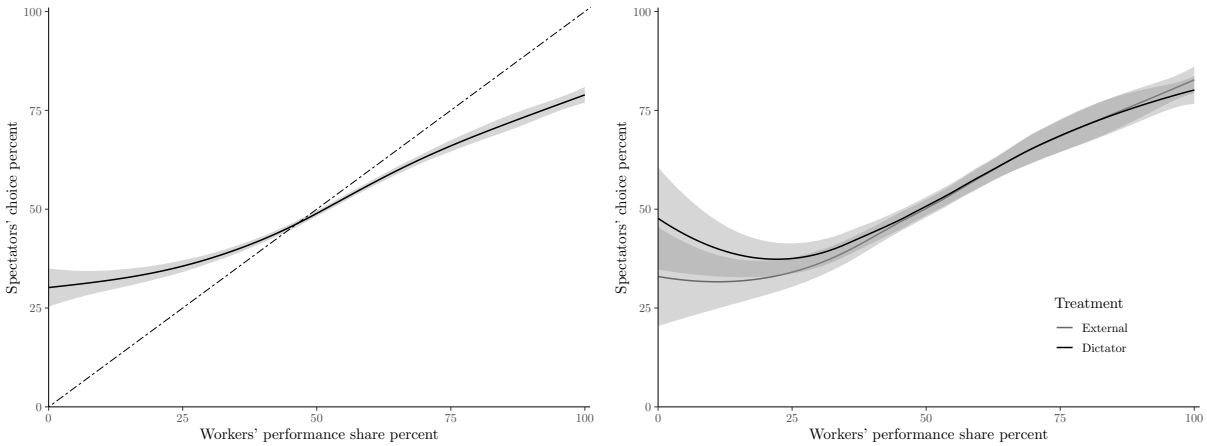
In this section, we first summarize general trends in spectator behavior before moving to the main analysis of the treatment effect. We then explore whether spectator responses vary depending on the type of allocation chosen by the worker, asking whether spectators are process-oriented or care about the dictator worker's perceived intent. Finally, we explore heterogeneity in spectator behavior.

4.1 Spectator decision-making

Figure 5 shows that across all initial choices, spectators implement an earnings floor by seldom assigning wages below 25% to any worker. Conversely, few workers receive wage shares higher than 75%, even when workers' performance shares exceed 90%. Thus, workers do not receive a 1-to-1 payout for performance; rather, spectators reward performance at a rate of ≈ 0.6.

A large majority, 81%, of spectators propose final earnings different from the initial earnings they are presented, with 71% deviating more than 5 percentage points.[12] The share of spectators deviating from wages proposed by a third party or a worker is similar in the DD (80%) and the ED treatment (82%). Although spectators generally show low acceptance of initial wages, they are less likely to deviate from initial wages aligned with fairness norms such as meritocratic or egalitarian pay.

12. Spectators could choose the initial earnings on the slider but there was no default to keep initial earnings. Spectators used a slider that started randomly at 0% or 100% and had to move the slider to proceed



(a) All initial choices (b) Meritocratic initial choices only

Figure 5: Mean spectators' choice by workers' performance.

Notes: Figures show the smoothed mean function with 95% confidence intervals. The dashed 45-degree line serves as a reference for wage shares equal to the performance share.

We call choices in which the initial wage proposal equals the performance share of the proposer, i.e., a worker producing 60%, getting 60% of the total earnings, "meritocratic" choices. Taking into account that most spectators choose to round to the nearest 5%, spectators accept 78% of all rounded meritocratic choices. This is well in line with observations from our previously-mentioned pilot study where 74 % of spectators held a meritocratic ideal point.

Figure 5 illustrates that spectators treat meritocratic choices similarly across treatment groups. The mean earnings in those cases are 57% in both groups. Note that as very few workers with a low performance (and by construction their third-party counterparts) opted for meritocratic initial pay, the confidence interval is large for choices below 25%.

In contrast to the relatively high acceptance rates of meritocratic pay, spectators generally make changes when initial earnings exceed performance shares with only 10% leaving earnings unchanged and 16% making changes smaller than 5 percentage points. Only 5% of spectators accept that 100% of earnings go to the dictator worker or their counterpart.

Having established general patterns of spectator behavior, we turn to analyzing differ-

ences in behavior across the DD and ED treatment. We follow the analysis set out in our pre-plan and move on to investigate in more detail when and why spectators react differently to wages chosen by a self-interested dictator worker compared to externally decided wages.

4.2 Main treatment effect

We start by testing our pre-registered main hypothesis: that mean final earnings in the DD treatment are lower than mean final earnings in the ED treatment. Spectators in the DD treatment may disapprove of workers choosing their own wage in principle and/or judge particular allocations as power abuse on the part of the workers. We compare how much spectators give the dictator worker choosing their own earnings in the DD treatment (mean final earnings), to how much spectators in the ED treatment give the corresponding worker with the same performance and initial earnings share. We present estimates from the following regression in the third column in Table 2.

$$w_c = \alpha_c + \beta X_{ci} + \gamma T_i + u_i \tag{1}$$

The outcome variable, w_c, is the wage share that spectators assign to the dictator in the DD group and their counterpart in the ED group. We measure the final earnings at the spectator-choice level in percentages. T is the binary treatment indicator variable, taking the value 1 when a spectator sees initial allocations proposed by the dictator worker (DD treatment) and the value 0 otherwise. γ is the parameter of interest that measures the difference in final wage shares between the DD and ED treatment. We expect a negative γ, reflecting lower average wages in the DD group. $\mathbf{X}$ is our matrix of control variables, and includes time- and batch dummies[13], as well as the performance share.

The first column in Table 2 reports the raw difference between the DD and ED treat-

13. We pre-registered to add dummies for batches of data collection, so that we could run additional rounds of data collection in case spectators dropped out during the experiment. Running additional rounds meant that we could collect a wage decision for all worker pairs in our sample.

ment.[14] We find that spectators assign 4 percentage points lower final earnings to dictator workers compared to their counterpart. The results do not change when controlling for performance (column 2) or when adding batch and hour controls (column 3).

The four percentage points lower earnings in the treatment group corresponds to 0.22 standard deviations of earnings. The average difference in final earnings between the DD and ED treatment results solely from who is deciding wages, as both initial choices and performances were identically distributed across the two groups.[15]

4.3 Treatment effect drivers: perceived intent

We pre-specified to interact treatment status choosing overproportional or underproportional pay to check what drives the average treatment effect.

The specification below includes the interactions between the DD treatment and two dummy variables indicating whether initial chosen earnings lie above or below the performance share. The reference category for both interactions is the earnings share assigned when performance and chosen earnings coincide (meritocratic choices). The interaction effect between treatment status and underproportional choices picks up spectators' assessment of what we call generous choices.

$$\begin{aligned} w_c = \alpha + \beta X_{ci} + \gamma T_i + \delta T_i * O_c + \theta T_i * U_c \\ + \pi O_c + \nu U_c + u_i \end{aligned} \tag{2}$$

O is the indicator variable for overproportional choices, while U is the indicator variable

14. In Appendix Table B1 we show that the main results are robust to restricting the sample to only spectators' first decision.

15. We measure fairness perceptions of our spectators as their willingness to redistribute earnings between the workers vis-à-vis the initial allocation. In order to pick up fairness perceptions in this way, spectators need to be willing to intervene. Ambuehl, Bernheim, and Ockenfels (2021) demonstrates that spectators are willing to intervene on behalf of someone else to stimulate more patient behavior, while Ackfeld and Ockenfels (2021) show at least limited willingness to intervene in order to induce more pro-social behavior. However, Charité et al. 2022 find that spectators respect recipients' reference points when redistributing endowment. By requiring spectators to intervene actively through redistribution, we may measure a lower bound of spectators' objections to power abuse.

for underproportional choices. δ and θ are the coefficient of interests for the interactions between the DD treatment group and the proportionality indicators.

Table 2: Average differences in assigned wage share between the DD and ED treatment

DV: Final wage share assigned by spectator				
	(1)	(2)	(3)	(4)
DD Treatment	−3.99	−3.99	−3.99	0.35
	(0.62)	(0.51)	(0.51)	(1.11)
Performance		0.60	0.60	0.59
		(0.02)	(0.02)	(0.02)
Overproportional				1.00
				(0.82)
Underproportional				−0.61
				(1.00)
DD Treatment x Overproportional				−5.75
				(1.30)
DD Treatment x Underproportional				−2.76
				(1.49)
Time Controls	No	No	Yes	Yes
Batch Controls	No	No	Yes	Yes
R^2	0.01	0.36	0.36	0.36
Adj. R^2	0.01	0.36	0.36	0.36
Num. obs.	3924	3924	3924	3924

Notes: In this table column (1) shows the raw difference in final assigned wages between the DD and ED treatment group without any control variables, in column (2) we control for performance and in column (3) we add a set of dummies for time of entry and batches. In column (4) overproportional is a dummy variable equal to 1 when the initial choice exceeds performance and similarly underporportional is a dummy variable equal to 1 when performance exceeds the initial choice. All control variables are specified in our pre-plan. All standard errors are clustered at the individual level.

The fourth column in Table 2 shows a large negative γ, reflecting that spectators react particularly to overproportional choices. Relative to meritocratic choices, worker dictators receive lower final wages when they get to pick their initial wages and make a self-serving choice. This difference is close to 6 percentage points which corresponds to 0.34 of a standard deviation. When accounting for this choice type, the baseline treatment term is small and insignificant, ruling out process concerns where spectators object to the mere fact that potentially self-interested workers make choices.

Spectators react more to substantially overproportional choices than to mildly overproportional choices. Figure 6 shows that choices that exceed the performance share by more than 25 percentage ppoints result in a larger penalty from the spectators, measured as the difference between the initial choice of the workers and the final choices of the spectators. For choices exceeding performance with less than 25 percentage points, both treatment and control group workers assign meritocratic earnings on average. Throughout, the 45-degree stippled line in Figure 6 corresponds to choices where spectators exactly correct an overproportional choice, assigning as much less as the initial wage exceeded the performance share. Anything above the 45-degree line indicates spectators assigning workers wages below performance share.

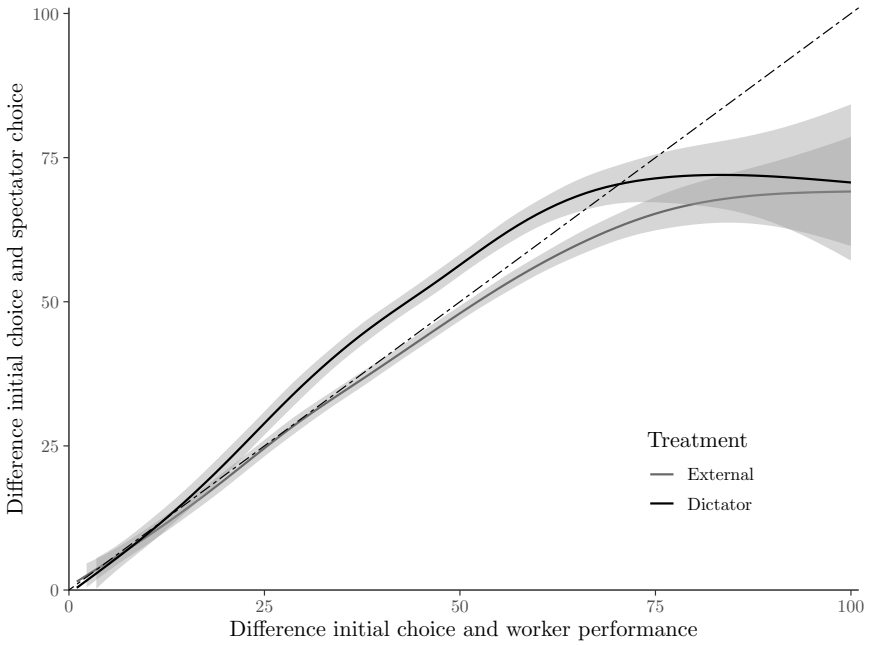


Figure 6: Spectators' reactions to overproportional choices.

Notes: This figure shows how spectators in the DD and ED treatment react differently to initial choices exceeding the performance share. The solid lines show the smoothed mean function with 95% confidence intervals. The sample is restricted to overproportional initial choices.

The final wages for workers making underproportional choices do not differ significantly between the treatment and control groups. If anything, workers receive lower final wages

when offering a generous wage share to their coworker than if the wages are suggested by a third party. Spectators may think that it is workers' right to give away some of their money (as indicated by their performance share), while the same choice seems unfair when made by a third party. Most of the variation in final wages for underproportional worker choices comes from workers who suggest egalitarian pay. Spectators tend to accept this if the higher-performing worker suggests it, while control group spectators tend to move the final wage share closer to the performance share when a third party suggests an egalitarian split. The answers in free text questions suggest that spectators in the DD treatment want to honor generous decisions of high-performing workers and consider the dictator worker entitled to share equally with their coworker in spite of superior performance.[16]

Free-text input by spectators indicates that spectators make value judgments about workers' motives, with no indications of a specific process concern. Figure 7 shows the words spectators used most frequently when asked to verbally reason their choices, for both treatment conditions. Spectators in the DD treatment seem to be concerned about worker dictators choosing high earnings for themselves. They use morally loaded words such as "greedy", "selfish" or "unfair" relatively more often and mention that they want to "punish" worker dictators or "teach them a lesson". Spectators in the external decision treatment use neutral vocabulary and prominently mention "time", "productive", "effort" and "deserve". Appendix Table B3 lists the words from the word cloud with frequencies and relative frequencies: While 60 people mentioned "selfish" in the DD treatment, spectators mentioned neither "selfish" nor "greedy" in the ED treatment. In contrast, "effort" was mentioned 102 times by spectators who saw external decisions compared to 58 times by spectators seeing workers' choices.

In our final pre-specified estimation equation, we interact treatment and a dummy vari-

16. In answers in free text questions asking spectators to give a reason for their choices, spectators mention that they perceived high-performer egalitarian choices as "generous" or "kind" and that they want to honor or respect those choices.

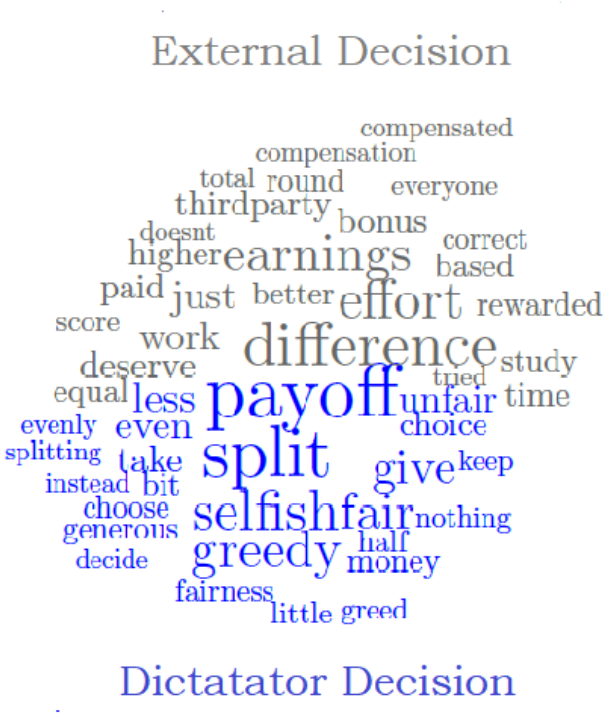


Figure 7: Word cloud from open-text question: Words used relatively more often by treatment condition.

Notes: After deciding the final allocations, spectators reasoned their choices in an open text question. The figure shows words used relatively more frequently in one treatment group compared to the other.

able for being the lower performer in the pair. We find that decision-making power reduces wages less for low performers than for high performers where we define high performers as a performance share exceeding 50%, see Appendix B.1. In the data, many low performers that make choices above performance share choose an egalitarian allocation. This pattern is in line with a broad definition of power abuse where spectators accept choices in line with an established fairness norm, even though this may not be their own fairness norm. Since low performing spectators that choose earnings above their performance share can land on egalitarian pay as a fairness-norm-supported allocation, this could explain why high performers exceeding their performance share fare worse than low performers exceeding their performance share.

4.4 Spectators' acceptance of fairness norms

Our results so far support *perceived-intent* rather than *process-oriented* spectators. Spectators do not generally assign lower wages to workers that choose their own wage share; rather, they decide on the basis of workers' concrete choices. In our pilot we showed that participants on Prolific overwhelmingly choose meritocratic allocations when not given any benchmark. We move on to ask whether spectators hold a narrow perceived-intent view of power abuse, where they accept only choices in line with their own fairness norm, or whether they tolerate a broader set of decisions, as long as they are in line with some widely recognized fairness norm.

To address this question, we examine whether the treatment effect varies according to which fairness norm the workers' initial allocation most closely reflects. We classify each initial choice into one of four categories: egalitarian, meritocratic, tournament pay, or evident power abuse. We label a choice as egalitarian if the worker either divides earnings evenly between themselves and their coworker or, in rare cases (48 total), if they performed below 50% of the output yet allocate themselves somewhere between their performance share and 50%. Most of these "partial" egalitarians choose 40% for themselves. A choice is deemed

meritocratic if the worker assigns themselves an amount equal to their performance share or slightly below it.[17] The remaining initial choices fall into two groups. If a high-performing worker assigns themselves more than their performance share, we label it as tournament pay. Conversely, if a worker who performed below 50% claims more than half of the total earnings, we classify it as evident power abuse since no existing fairness norm condones a low performer taking more than half.

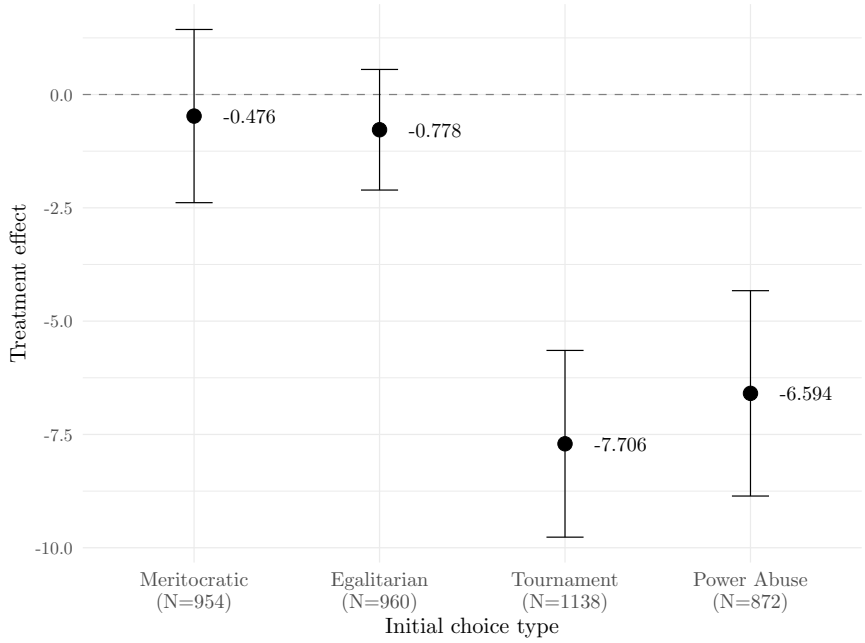


Figure 8: Treatment effects by choice type.

Notes: The plot shows the estimated treatment effect (γ) from Equation (1) separately for egalitarian, meritocratic, tournament pay, and power abuse choices. Regressions include controls.

Figure 8 shows the treatment effect by choice categories. For both meritocratic and egalitarian allocations, the difference between treatment and control is close to zero and insignificant. By contrast, for power abuse and tournament pay initial choices, the treatment effect is -6.6 and -7.7, respectively. We interpret these findings to mean that spectators

17. We include these marginal "under-allocations" in the meritocratic category to retain our full sample. All results remain robust if we apply stricter definitions of meritocracy and egalitarianism and remove cases that cannot be categorized according to stricter definitions from the sample. See Appendix Figure B1 with strict definitions of egalitarian choices as exactly 50% and meritocratic chioces as exactly equal to performance.

readily accept stakeholders' meritocratic and egalitarian choices, but they respond negatively to tournament-style or power-abuse allocations when proposed by a self-interested party.

These results speak directly to the question of whether spectators adopt a broad or narrow definition of power abuse when they judge initial allocations. Our finding that neither egalitarian nor meritocratic allocations generate a penalty shows that spectators do not seem to narrowly impose their own fairness norm.

In our sample, workers that choose egalitarian allocations tend to be low performers and workers that choose meritocratic allocations tend to be high performers. They choose whichever fairness norms suits them best without simply claiming everything, striking a balance between self-interest and social desirability. This aligns with moral wriggling observed in Konow (2000), Cappelen et al. (2007), and Dana, Weber, and Kuang (2007).

Our findings suggest that spectators give these workers the benefit of the doubt - after all, workers make only one decision, and spectators cannot know for sure whether they would have chosen an egalitarian share also if they had performed better. Instead, they reserve their contempt for workers whose choices clearly violate established fairness norms. Spectators penalize tournament pay, a higher performer taking everything, just as much as evident power abuse where the lower performer takes more than their share. This may be related to the explicit framing of the effort task in terms of production in a pair, rather than a competition in which one plays to win.

5 Exploratory analysis

We turn our focus to overproportional choices, bundling the equally rejected tournament pay and power abuse choices, and describe how control and treatment group spectators react to these choices.

On average, spectators in the DD treatment assign lower wages when workers claim overproportional pay. We interpret this penalty as punishment for power abuse. In this

section, we identify under what conditions spectators assign workers earnings below their performance share. Next, we move on to explore which groups of spectators engage in punishment, using information on socio-economic background and political orientation from our exit survey.

5.1 Punishment

We use the following definition of punishing choices: Whenever the share of initial earnings exceeds the performance share and a spectator reacts by assigning final earnings below the performance share ($final\ wage < performance < initial\ wage$), we refer to this as "punishment". In contrast, we refer to spectators who do not accept overproportional pay, but simply revert to performance pay ($final\ wage = performance < initial\ wage$), as "correcters" and those accepting overproportional pay as "accepters".

Control group spectators serve as the benchmark to detect the treatment effect. Broadly, control group spectators *correct* wages back to the performance shares. For the group of choices where a third party proposes earnings above the performance share (ED), spectators in the control group assign on average final earnings of 51.2%, very close to the average performance of 50%.

The share of correcting and punishing decisions for spectators differs notably across the DD and ED treatment: In the DD treatment, 35% of spectators decide to punish dictator workers choosing overproportional wage shares. The remaining two-thirds of DD spectators divide almost equally between correcting the choices (down) to the meritocratic share and accepting a share that exceeds performance.

In contrast, only 23% of spectators in the ED treatment make punishing choices and a correspondingly higher share corrects earnings exactly to the meritocratic share. The difference between the share of punishing choices across the two treatment conditions has a t-value of 6.9.

In fact, the increased share of spectators making "punishing" choices in the DD treatment

come mostly at the expense of "correcting" choices. While 40 % of spectators make correcting choices in the ED treatment, only 31 % do so in the DD treatment (t-value of 4.6). Thus, it seems as if those moved by our treatment are spectators who punish workers who choose overproportional pay for themselves, while they would correct rather than punish, had the same allocation been proposed by a third party. A majority of "accepters" (80% in the DD and 70% in the ED treatment) are spectators facing relatively worse performing workers whose initial wages exceed their performance share. In these cases, the average "accepting" spectator allows for an overproportional share to the worse performing worker by assigning an equal split of earnings.

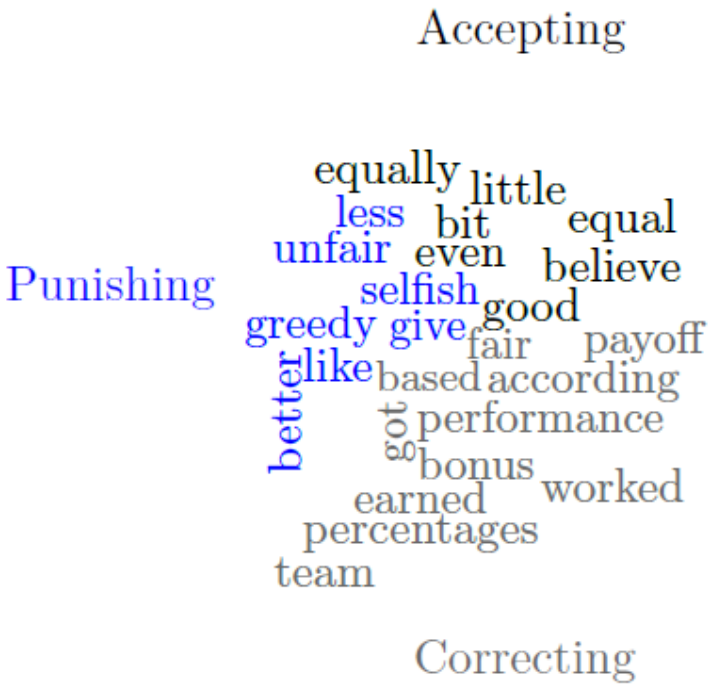


Figure 9: Word cloud comparison of "punishers", "correcters" and "accepters" in the DD treatment.

Notes: The word cloud shows words relatively more frequently used by spectators who made punishing, accepting, or correcting choices in the DD treatment for overproportional initial choices.

Figure 9 shows the wordcloud of words more frequently used by "punishers" compared to "accepters" and "correcters" in the DD treatment. "Accepters" mention "equally" more

often than the other two groups, possibly capturing egalitarian motives. "Correcters" use words like "performance" and "percentage" relatively more often, indicating that they perceive performance as the most important factor determining which wages workers should get. For the "punishers", words like "greed" and "selfishness" occur relatively frequently, again illustrating that these spectators perceive overproportional choices as self-serving and therefore assign low wages to the dictator. We show in Appendix Table B4 that when controlling for performance and initial wage, mentioning "greed/selfishness" is correlated with punishing, while mentioning "performance" is negatively correlated with making a punishing choice in the DD treatment.

The extra 12 percentage points of spectators deciding to punish in the DD treatment compared to the ED choice treatment may seem low. However, this comparison only picks up on the prevalence of any punishing choice, the extensive margin of the treatment effect. Turning to the intensive margin, we see that the average spectator in the DD treatment punishes by 19.5 percentage points, compared to 10.4 percentage points in the ED treatment group. On average, spectators punish almost twice as hard when faced with a dictator worker making a self-serving choice.

Figure 10 shows the distribution of final earnings between the DD and ED choice treatment for all punishing choices. In the ED treatment, reverting to egalitarian pay seems to be the most common response in cases where a third party proposes overproportional pay for a worker. For the DD treatment, the picture is very different. A sizable proportion of spectators decides to assign the lowest possible final wage share - 0% - which is a very uncommon choice for ED treatment spectators. While we see that those who punish, tend to do so harshly, the remaining 2/3 of spectators in the DD treatment make choices similar to those in the ED treatment.

In this section we have shown that we can trace the difference in final wages across the two treatment conditions to spectators that correct overproportional wages back to the per-

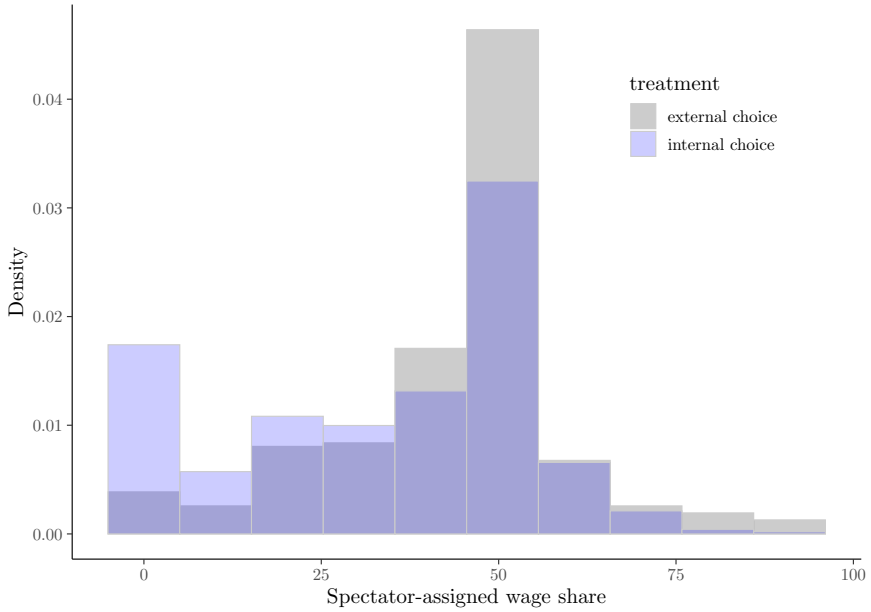


Figure 10: Wage bunching in punishment.

Notes: This figure shows the distribution of spectator choices in cases with *overproportional* initial choices and *punishing* spectator choices, i.e., initial wage share > performance share *and* final wage share < performance share.

formance share when wages are decided externally but instead punish when dictator workers choosing their own wage. In the next section, we investigate the characteristics of spectators switching from being "correcters" in one treatment condition, to being "punishers", in another.

5.2 Heterogeneity

Table 3 shows how the interaction between treatments and i) political orientation[18], ii) believing that justice will prevail and iii) self-positioning on a social ladder, predicts punishing and correcting behavior. The results indicate that in the DD treatment those punishing are more left-winged/liberal, and those correcting more right-winged/conservative (only left-right in the following) than in the ED treatment. Note that the two treatment conditions are balanced with respect to political views as shown in Table 1. The variation in the

18. In the US we measured political orientation on a scale from liberal to conservative, in Europe we measured political orientation on a left-right scale

inclination to make punishing choices also affects the overall treatment effect, with high-SES and left-leaning spectators driving the effects. Appendix Table B5 shows that interacting the treatment-dummy with being left-winged and believing that justice will prevail halves the baseline treatment effect, although the interaction-term itself is not significant. Our measure of trust, which is believing that most people would "try to be fair" is correlated with having a smaller treatment effect and the interaction is significant at the 5% level. This result indicates that the treatment effect is 25% larger comparing only spectators believing that "most people would take advantage of you", than it is when comparing spectators believing that "most people would try to be fair".

Table 3: Table showing how the interaction between DD treatment and answers to survey questions predict punishing or correcting behaviour

	(1) Punishment1	(2) Correcting1	(3) Punishment2	(4) Correcting2	(5) Punishment3	(6) Correcting3
DD Treatment	0.23	-0.16	0.03	0.08	0.01	-0.06
	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
Left-winged	0.02	-0.02				
	(0.01)	(0.02)				
DD Treatment x Left-winged	-0.05	0.04				
	(0.02)	(0.02)				
Confident justice prevails over injustice			0.00	0.03		
			(0.01)	(0.01)		
DD Treatment x Justice prevails overinjustice			0.03	-0.05		
			(0.02)	(0.02)		
Place on Socio-economic ladder					0.00	0.00
					(0.01)	(0.01)
DD treatment x Socio-economic ladder					0.02	-0.00
					(0.01)	(0.01)
R^2	0.11	0.01	0.11	0.01	0.11	0.01
Adj. R^2	0.10	0.01	0.10	0.01	0.10	0.01
Num. obs.	2565	2565	2588	2588	2587	2587

Notes: Columns 1,3 and 5 present coefficients from a regression with *punishment*, a dummy equal to 1 for cases where $final\ wage < performance < initial\ wage$ and 0 otherwise, as the dependent variable. Column 2,4 and 6 present coefficients from a regression with *correcting*, a dummy equal to 1 for cases where $final\ wage = performance < initial\ wage$ and 0 otherwise, as the dependent variable. All regressions include only spectators facing overproportional initial choices and control for performance and initial choice. *Left-winged* refers to a dummy variable equal to 1 for individuals who in our survey respond being "left"/"center-left" and 0 otherwise. *Confident justice prevails over injustice* refers to answer to a question on whether justice will prevail over injustice on a scale of 5 from "strongly agree" to "strongly disagree". *Place socio-economic ladder* refers to the respondent's self-placement on a ladder from 1 to 10 where the top are those with the most money, best jobs etc. The full questions are listed in Appendix A.3. In addition to controlling for performance, all regressions also include time and batch controls. Standard errors are clustered at the individual level.

6 Conclusion

In this paper, we present the results of a large online experiment designed to explore how people react to income inequality arising from discretionary decision-making power. The experiment consisted of two stages. First, we recruited a group of participants, referred to as "workers", who performed a task and were paired with one another. Each pair received a bonus to be divided between them. We randomized whether the split of the bonus, referred to as initial earnings, was determined by one of the workers in the pair (the "dictator") or by an external third party. Importantly, we designed the experiment so that the distribution of earnings across these two groups was identical. In the second stage, we recruited a separate group of participants, impartial spectators, to decide on final earnings. Spectators were randomly assigned to one of two conditions: in the Dictator Decides (DD) treatment, spectators evaluated allocations made by workers deciding their own earnings; in the External Decision (ED) treatment, they evaluated earnings determined by a third party.

We find that decision-making power decreases inequality acceptance, but only when power is perceived as abused. In our conceptual framework, we asked whether spectators are primarily process-oriented, objecting to stakeholders' decision-making power in principle, or whether they judge outcomes based on perceived intent. Our findings support the latter: spectators in the two treatments responded similarly to meritocratic and egalitarian allocations but reacted strongly when dictator workers awarded themselves earnings that exceeded their relative performance. This includes both low-performing workers taking more than 50% of the bonus and high-performing workers opting for "tournament pay" by assigning themselves disproportionately high rewards.

Moreover, we find that the difference between the ED and DD treatment is driven by a subset of spectators assigning earnings below performance to dictators who have chosen initial earnings exceeding their performance share. We refer to these spectators as "pun-

ishers". We explore background characteristics associated with punishing behavior and find that spectators who punish tend to be more left-winged, less trusting and believe that justice will prevail injustice.

We contribute to the literature on redistribution preferences by demonstrating that power abuse in the income-generating process undermines inequality acceptance. Spectators apply a nuanced fairness framework: they do not rigidly enforce their personal fairness ideals nor uniformly reject stakeholder decision-making. Rather, they respect a plurality of fairness norms, such as meritocracy and egalitarianism—and intervene only when a decision falls outside a collectively recognized "justifiable range."

Although we acknowledge that we have removed many important features of inequality in society in our experiment, we believe that our findings contribute to understanding the variation in responses to inequality levels. Our findings suggest that people may accept less inequality when they perceive inequality to stem from power abuse rather than being externally determined.

In the present study, spectators have perfect information on whether worker dictators abused their decision-making power for their own benefit. In real-world income processes observers can only draw on signals about whether power abuse has taken place. An interesting venue for future research would be to explore whether spectators also redistribute allocations in a setting that includes uncertainty about power abuse.

References

Ackfeld, Viola, and Axel Ockenfels. 2021. "Do People Intervene to Make Others Behave Prosocially?" *Games and Economic Behavior* 128:58–72.

Akbaş, Merve, Dan Ariely, and Sevgi Yuksel. 2019. "When is inequality fair? An experiment on the effect of procedural justice and agency." *Journal of Economic Behavior & Organization* 161:114–127.

Alesina, Alberto, and George-Marios Angeletos. 2005. "Fairness and Redistribution." *American Economic Review* 95 (4): 960–980.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Preferences for Redistribution in the Land of Opportunities." *Journal of Public Economics* 89 (5-6): 897–931.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational Mobility and Preferences for Redistribution." *American Economic Review* 108 (2): 521–54.

Almås, Ingvild, Alexander Cappelen, and Bertil Tungodden. 2020. "Cutthroat Capitalism Versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy* 128 (5): 1753–1788.

Ambuehl, Sandro, B Douglas Bernheim, and Axel Ockenfels. 2021. "What Motivates Paternalism? An Experimental Study." *American Economic Review* 111 (3): 787–830.

Arechar, Antonio A., Simon Gächter, and Lucas Molleman. 2018. "Conducting Interactive Experiments Online." *Experimental Economics* 21 (1): 99–131.

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2022. "Labor Market Concentration." *Journal of Human Resources* 57 (S): S167–S199.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor Market Power." *American Economic Review* 112 (4): 1147–93.

Bortolotti, Stefania, Ivan Soraperra, Matthias Sutter, and Claudia Zoller. 2023. "Too Lucky to Be True: Fairness Views Under the Shadow of Cheating." *Review of Economics and Statistics,* 1–45.

Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. 2007. "The Pluralism of Fairness Ideals: An Experimental Approach." *American Economic Review* 97 (3): 818–827.

Cappelen, Alexander W, Karl Ove Moene, Siv-Elisabeth Skjelbred, and Bertil Tungodden. 2023. "The Merit Primacy Effect." *The Economic Journal* 133 (651): 951–970.

Card, David. 2022. "Who set your wage?" *American Economic Review* 112 (4): 1075–90.

Charité, Jimmy, Raymond Fisman, Ilyana Kuziemko, and Kewei Zhang. 2022. "Reference Points and Redistributive Preferences: Experimental Evidence." *Journal of Public Economics* 216:104761.

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

Edlund, Jonas. 1999. "Trust in Government and Welfare Regimes: Attitudes to Redistribution and Financial Cheating in the USA and Norway." *European Journal of Political Research* 35 (3): 341–370.

Fehr, Ernst, Thomas Epper, and Julien Senn. 2024. "Social Preferences and Redistributive Politics." *Review of Economics and Statistics,* 1–45.

Fehr, Ernst, Urs Fischbacher, and Simon Gächter. 2002. "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms." *Human Nature* 13:1–25.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415 (6868): 137.

Gärtner, Manja, Johanna Mollerstrom, and David Seim. 2017. "Individual Risk Preferences and the Demand for Redistribution." *Journal of Public Economics* 153:49–55.

Hvidberg, Kristoffer B, Claus T Kreiner, and Stefanie Stantcheva. 2023. "Social Positions and Fairness Views on Inequality." *Review of Economic Studies* 90 (6): 3083–3118.

Kleven, Henrik Jacobsen. 2014. "How Can Scandinavians Tax So Much?" *Journal of Economic Perspectives* 28 (4): 77–98.

Klimm, Felix. 2019. "Suspicious success–Cheating, inequality acceptance, and political preferences." *European Economic Review* 117:36–55.

Knutsen, Tora, and Sonja Kovacevic. 2021a. *An Experiment on How Wage Discretion Affects Distribution Preferences.* AEA RCT Registry, October. https://doi.org/10.1257/rct.8291-3.0.

———. 2021b. *To Give or Not to Give. Explaining Earnings Choices in an Online Labour Market Experiment.* AEA RCT Registry, October. https://doi.org/10.1257/rct.8345-1.0.

Konow, James. 2000. "Fair shares: Accountability and cognitive dissonance in allocation decisions." *American Economic Review* 90 (4): 1072–1091.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments." *American Economic Review* 105 (4): 1478–1508.

Litman, Leib, Jonathan Robinson, and Cheskie Rosenzweig. 2015. "The Relationship Between Motivation, Monetary Compensation, and Data Quality Among US-and India-Based Workers on Mechanical Turk." *Behavior Research Methods* 47 (2): 519–528.

Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø Sørensen. 2015. "Luck, Choice and Responsibility: An Experimental Study of Fairness Views." *Journal of Public Economics* 131:33–40.

Palan, Stefan, and Christian Schitter. 2018. "Prolific.ac-A Subject Pool for Online Experiments." *Journal of Behavioral and Experimental Finance* 17:22–27.

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?" *The Quarterly Journal of Economics* 136 (4): 2309–2369.

A Details experiment

A.1 Instructions

Description of study shown to potential participants in Prolific before they decided to participate in the study:

> *"You will decide payments to participants of an experiment conducted on Prolific between October 12th and18th.*
>
> *This is not a hypothetical survey. Please make sure that you make a careful decision. It is crucial that you read all instructions carefully and understand all elements of the study. We ask you to only decide to take this study if you feel comfortable doing a lot of careful reading in English. The study contains an open text question that needs to be answered in English.*
>
> *Careful: To participate in this experiment, you should use either google chrome or mozilla firefox. The experiment does not work well on internet explorer! Make sure you fill in your ProlificID in the start, otherwise we do not know who to pay the participation fee."*

A.2 Screenshots

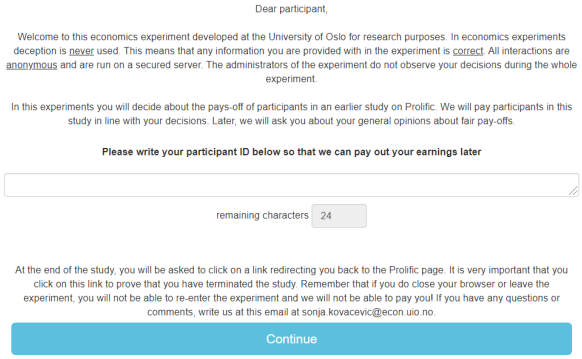
A.3 Background questions

– **Q1: Wage deservingness** To what extent to you think those working in the following occupations get paid what they deserve? (1.Much less than they deserve ... 5.Much more than they deserve) [We will randomize whether participants get this question or question 6 here]

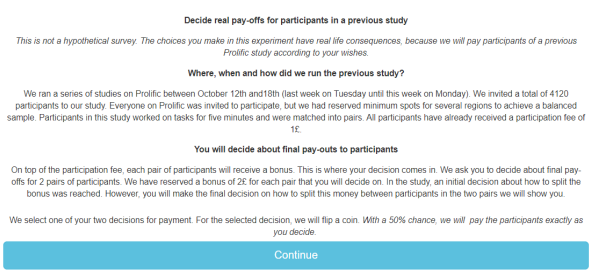
Occupations:

1. Politician

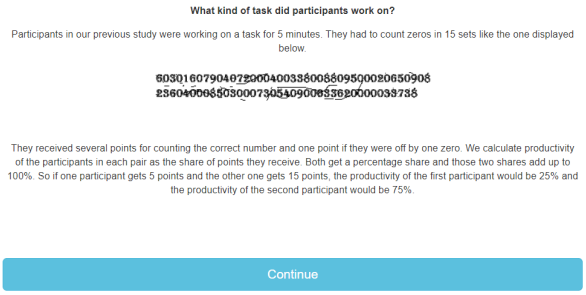
Figure 11: Information to spectators (same for all).



(a) Introduction to experiment

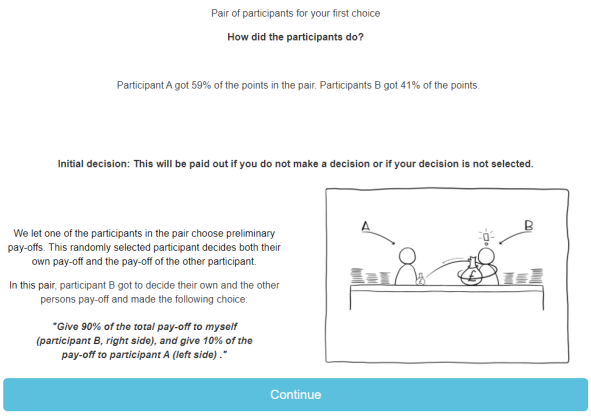


(b) Introduction of workers

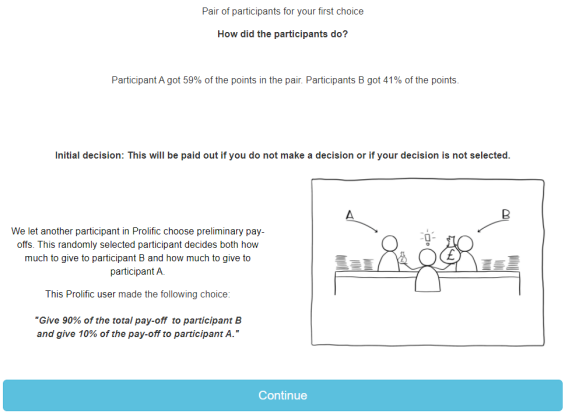


(c) Information about the task workers did

Figure 12: Information about initial earnings.

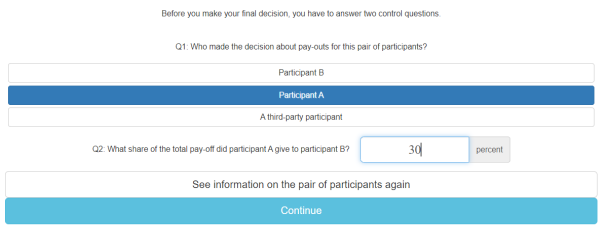


(a) DD treatment

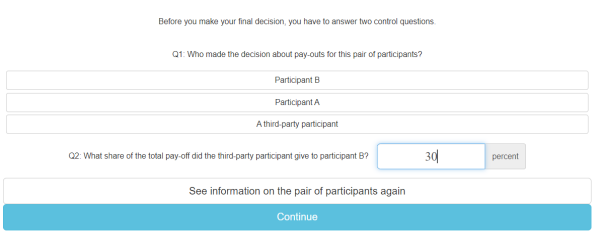


(b) ED treatment

Figure 13: Control questions.

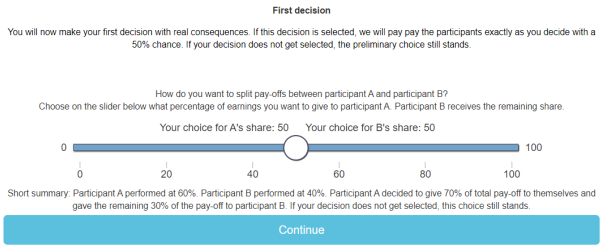


(a) DD treatment

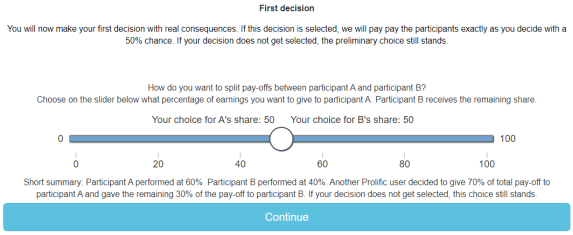


(b) ED treatment

Figure 14: Decision screen.

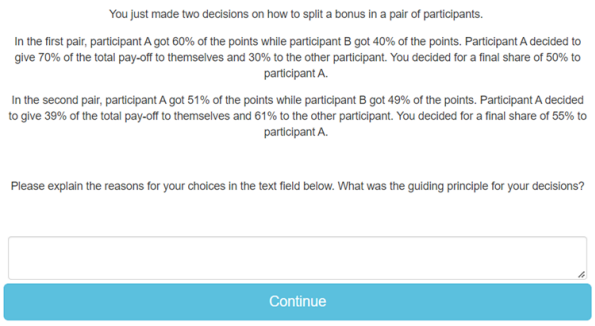


(a) DD treatment

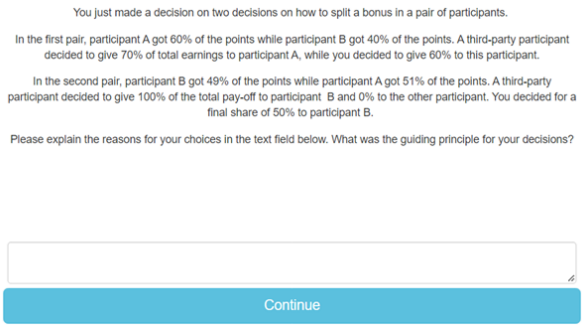


(b) ED treatment

Figure 15: Open-text question.



(a) DD treatment



(b) ED treatment

2. Football player
3. CEO of a large company
4. Taxi driver
5. Social worker

Value questions:

In the following, we will ask you to place your views on the scale below. 1 means you agree completely with the statement on the left; 10 means you agree completely with the statement on the right. If your views fall somewhere in between, you can choose any number in between.

- **Q2:** People can only get rich at the expense of others 1.....10 Wealth can grow so there's enough for everyone.
- **Q3:** There should be greater incentives for individual effort- 1.....10- Incomes should be made more equal
- **Q4:** Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair? [Most people would try to take advantage of me 1 ..... 10 most people would try to be fair]
- **Q5:** I am confident that justice always prevails over injustice [Agree strongly Agree Neither agree nor disagree Disagree Disagree strongly]

Beliefs wage discretion

- **Q6:** To what extent to you think can people in the following occupations decide their own pay? (1.They cannot decide their own pay... 5.They can decide their own pay) [We will randomize whether participants get this question or question 1 here. Occupations the same as in question 1]

- **Q7 Not from the U.S:** In political matters, people talk of "the left" and "the right." How would you place your views on this scale, generally speaking?

    1. Left
    2. Center left
    3. Center right
    4. Right

- **Q7: Political views, from the U.S** In political matters, people talk of "Liberals" vs. "Conservatives". Which of these positions corresponds most closely to your views.

    1. Liberal; 2 Slightly liberal; 3 Slightly conservative; 4 Conservative

- **Q7.1 (only U.S.)** Who did you vote for in the last presidential election?

    1. Donald Trump
    2. Joe Biden
    3. Other
    4. Did not vote

- **Q8: Social ladder**

    Think of a ladder (see image) as representing where people stand in society. At the top of the ladder are the people who are best off—those who have the most money, most education and the best jobs. At the bottom are the people who are worst off who have the least money, least education and the worst jobs or no job. The higher up you are on this ladder, the closer you are to people at the very top and the lower you are, the closer you are to the bottom. Where would you put yourself on the ladder? Choose the number whose position best represents where you would be on this ladder

B Analysis

B.1 Additional Pre-specified estimations

We pre-registered to check whether the main results hold if we look only at spectators' first decision. These results are shown Table B1. We also pre-registered to check whether the treatment effect is different when it is the low-performing worker who got to decide initial wages. This means that we run the following regression:

$$w_c = \alpha + \gamma T_i + \delta LowPerformer_c * T_i + \theta LowPerformer_c + \beta X_{ci} + u_i \tag{3}$$

Where $LowPerformer$ is an indicator equal to 1 if the preliminary earnings belong to the lower performer in the worker pair. This means that for the treatment group these are low-performing workers who determined wages. δ then measures the difference in final wages between treatment and control group spectators in cases where they decided the share for the lower performing worker. As X still includes performance, δ measures the difference between the treatment and control group for the same performances. The results from this estimation are shown in Table B1.

B.2 Analysis open-text question

After spectators had made decisions on how to split earnings within the worker pairs, spectators had to explain their choices in an open-text question. The screenshots in Figure 15 show how the spectators were asked to give the reasons for their decisions.

In order to make an interesting comparison of words used by spectators in the two treatments, we first remove words that are frequent but provide little information. The R-package "stopwords" provides a list of 169 English words such as pronouns, prepositions and time adverbs. We also remove punctuation and reference to numbers. In addition we pre-specified removing some other words frequently used in the pilot that gave little

Table B1: Additional pre-registered regressions

(a) Main results using only spectators' first decisions

	(1)	(2)
DD treatment	−3.34	−0.53
	(0.63)	(1.61)
Performance	0.60	0.61
	(0.02)	(0.02)
Overproportional		1.01
		(1.28)
Underprop		−1.55
		(1.54)
DD treatment x Overproportional		−3.79
		(1.78)
DD treatment x Underproportional		−1.56
		(2.17)
R^2	0.40	0.40
Adj. R^2	0.39	0.39
Num. obs.	1962	1962

Notes: The left panel presents coefficients from regression 1 and 2 using only spectators' first decision. The right panel shows coefficients from estimating Regression 3. All regressions include controls.

(b) Treatment effect is smaller when low performers choose

	(1)	(2)
DD treatment	−5.30	−5.28
	(0.80)	(0.71)
Low performer	−18.56	−3.64
	(0.70)	(0.75)
DD treatment x Low performer	2.86	2.88
	(1.10)	(1.03)
Performance		0.55
		(0.03)
R^2	0.23	0.36
Adj. R^2	0.22	0.36
Num. obs.	3924	3924

information.[19]

Next, we make a frequency table by treatment group, this means that we count how many times for each word was used in the DD and ED treatment. We keep only those words that are used at least 10 times in one of the treatment groups. Finally we compute the relative frequency of a word in one treatment group compared to the other. Table B3 shows the top 20 used more frequently by spectators in one treatment group than in the other.

Table B3: Top 20 words used relatively more frequent in one treatment group compared to the other

(a) More frequent DD treatment

	ED	DD	Rel. Freq.
greed	0	16	Inf
selfish	0	60	Inf
honest	2	11	5.50
willing	2	11	5.50
instead	4	19	4.75
choose	6	26	4.33
partner	4	17	4.25
initially	3	12	4.00
splitting	5	19	3.80
payoff	36	130	3.61
decide	6	20	3.33
winnings	4	13	3.25
though	18	57	3.17
nothing	8	25	3.12
take	16	46	2.88
matter	6	17	2.83
agreed	5	14	2.80
agree	4	11	2.75
keep	10	27	2.70
unfair	20	54	2.70

(b) More frequent ED treatment

	ED	DD	Rel. Freq.
thirdparty	23	0	Inf
round	19	2	9.50
party	23	5	4.60
third	22	6	3.67
difference	81	23	3.52
productive	15	5	3.00
rewards	15	5	3.00
compensated	17	6	2.83
attention	11	4	2.75
earnings	72	28	2.57
harder	12	5	2.40
compensation	20	9	2.22
big	19	9	2.11
numbers	18	9	2.00
outcome	15	8	1.88
survey	11	6	1.83
effort	102	58	1.76
different	19	11	1.73
mean	12	7	1.71
accordingly	13	8	1.62

Notes: This table shows the frequencies of words used relatively more often in the DD vs. ED treatment. All words are from the open-text question.

19. These are: “actually”, “think”, “also” , “participant”, “participants”, “therefore”, “someone”, “option”, “thought”, “made”, “worker”, “workers”, “person”, “decided”, “amount”,“didnt”, “chose”,“still”, “can”,“put”, “one”, “first”, “second”, “last”, “final”, “etc”, “pair”, “isn’t”, “pairs”, “know”, “player”, “something”, “seems”, “may”, “pay”, “might”, “felt”, “thats”,“hence”, “will”, “cases”, “way”, “simply”,“used”, “main”.

Table B4: Correlation of word use with punishing choices

DV: Punishing	(1) All	(2) Within DD	(3) Within ED
DD Treatment	0.11		
	(0.02)		
Initial choice	0.00	0.01	0.00
	(0.00)	(0.00)	(0.00)
Performance	0.01	0.00	0.01
	(0.00)	(0.00)	(0.00)
Mentions greed	−0.11	0.29	−0.18
	(0.04)	(0.04)	(0.03)
Mentions performance	0.09	−0.06	0.09
	(0.03)	(0.03)	(0.03)
Mentions equal	0.13	0.07	0.13
	(0.04)	(0.04)	(0.04)
DD Treatment x mentions greed	0.41		
	(0.05)		
DD Treatment x mentions performance	−0.15		
	(0.05)		
DD Treatment x mentions equal	−0.07		
	(0.05)		
R^2	0.14	0.19	0.07
Adj. R^2	0.14	0.19	0.07
Num. obs.	2610	1305	1305

Notes: Sample includes only spectators facing overproportional initial choices. Errors clustered at the individual level.

Table B5: Treatment effect interacted with background variables

	(1)	(2)	(3)	(4)	(5)	(6)
DD Treatment	-2.70	-4.67	-4.02	-4.72	-6.05	-2.01
	(0.99)	(0.85)	(1.08)	(1.25)	(1.04)	(1.72)
Left-leaning	-0.15					
	(0.69)					
Performance	0.60	0.60	0.60	0.60	0.60	0.60
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
DD Treatment x Left-leaning	-1.80					
	(1.16)					
Female		0.23				
		(0.64)				
DD Treatment x Female		1.13				
		(1.05)				
Age			-0.03			
			(0.03)			
DD Treatment x Age			0.00			
			(0.03)			
Incomes should be more equal				0.02		
				(0.11)		
DD Treatment x Incomes should be more equal				0.12		
				(0.19)		
Most people would be fair					-0.16	
					(0.13)	
DD Treatment x Most people would be fair					0.47	
					(0.21)	
Confident justice prevails over injustice						-0.48
						(0.29)
DD Treatment x Justice prevails over injustice						-0.58
						(0.49)
Pr(>F)	0.033	0.13	0.98	0.46	0.014	0.22
R^2	0.36	0.36	0.36	0.36	0.36	0.36
Adj. R^2	0.36	0.36	0.36	0.36	0.36	0.36
Num. obs.	3856	3894	3880	3898	3888	3886

Notes: This table shows results from regressions where the treatment-dummy is interacted with background variables. Regressions include controls for performance, time, and batch. Standard errors are clustered at the individual level.

B.3 Participants by region

Prolific is available to users in OECD-countries and South Africa. We invited participants from frequently studied regions, such as the U.S., Scandinavia, and Western Europe to draw comparisons to the existing literature. This literature covers differences in redistribution preferences between (Western) Europe and the U.S. (e.g. Alesina and Angeletos 2005; Alesina and La Ferrara 2005; Alesina, Stantcheva, and Teso 2018) and Scandinavia and the U.S. (e.g. Almås, Cappelen, and Tungodden 2020; Kleven 2014; Edlund 1999). Additionally, we invited some participants from less-studied regions to study spectator responses across different institutional settings and inequality levels.

A pilot study revealed that participants from some areas entered our study faster than others. Therefore, we ran 7 studies simultaneously, each restricted to an area and with a fixed pre-specified share of total participants per area: a) South Africa (8.5%), b) Mexico (8%), c) United States (27%), d) United Kingdom (10%), e) Western Europe (22%), f) Scandinavia (8.5%), and g) Eastern and Southern Europe (15%).[20]

Figure B2 shows that the treatment effect is present for spectators from many countries, but not all: For spectators from Western Europe, the U.S., Scandinavia and Mexico, the difference between the ED and DD treatment is large and significant, while it is non-distinguishable from 0 for spectators from the U.K., South Africa and the rest of Europe.

20. In Prolific researchers can restrict for whom the study will be available. According to Prolific, they have the most participants in the U.K. and the U.S. However, either because fewer studies are available to them or because they are more active, respondents from South Africa and Mexico were the quickest to enter the study. Some of the groups, in particular Scandinavians, entered at such a slow rate that we had to run the study over several days. As we wanted hour/day fixed effects, we had to get participants from different groups on all days. We did this by slowly increasing the spots in the studies.

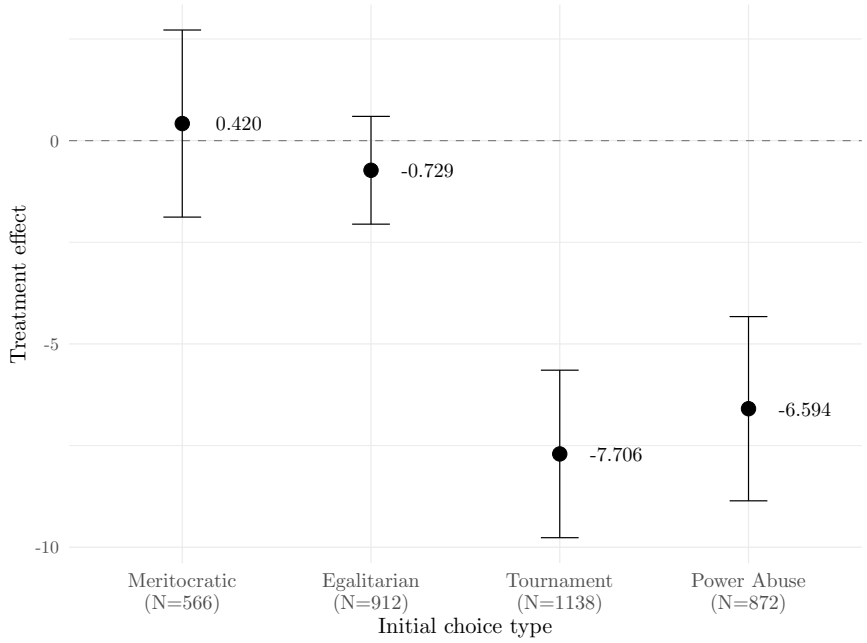


Figure B1: Treatment effects by choice type with strict definition of egalitarian and meritocratic choices.

Notes: The plot shows the estimated treatment effect for different choice types using strict definitions (e.g., egalitarian is exactly 50%). All regressions include controls.

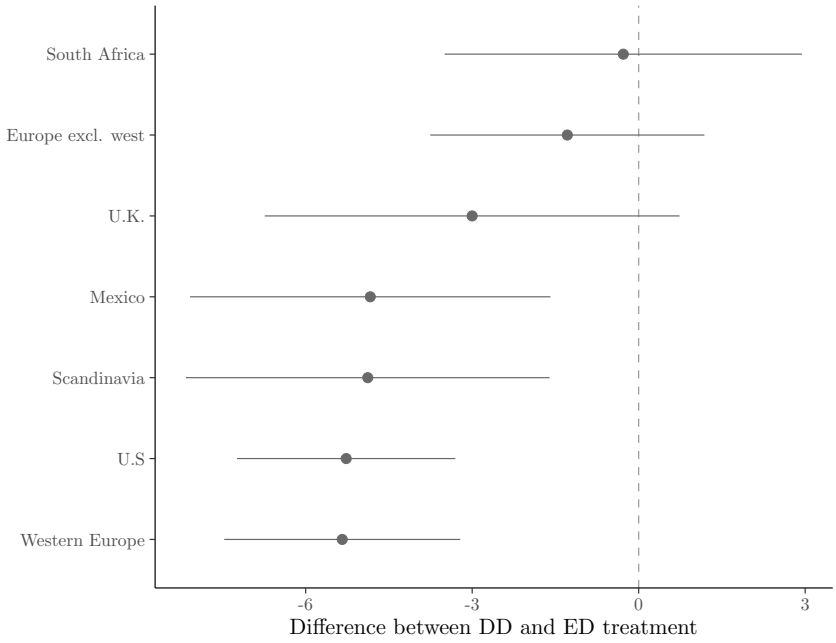


Figure B2: Average effect by country or region.

Notes: The plot shows the estimated treatment effect by country/group of countries. All regressions include a linear term controlling for performance and clustered standard errors.