Excessive pay is not about the numbers How power abuse erodes inequality acceptance^{*}

Tora Knutsen¹ and Sonja Kovacevic²

¹NIFU. t.k.knutsen@nifu.no ²University of Oslo. sonja.kovacevic@econ.uio.no

November 15, 2023

Abstract

We investigate how inequality acceptance depends on power abuse. Running a largescale pre-registered two-stage experiment, we measure inequality acceptance through spectators redistribution choices. We randomize whether a worker can decide their initial earnings, potentially abusing their power for their own gain at the expense of their co-worker by allocating high earnings to themselves. Then, impartial spectators redistribute. We find that spectators give significantly lower earnings to workers deciding their own initial earnings, compared to workers with identical but externally decided initial earnings. Spectators redistribute substantially when confronted with power abuse, while they accept meritocratic initial earnings regardless of who proposes. While most spectators redistribute to achieve meritocratic shares, a minority of spectators switches to active punishment in the face of power abuse. Thus, allowing for power abuse dismantles a consensus to implement meritocratic (and unequal) pay.

Keywords: Inequality; Fairness; Redistribution; Experiment.

JEL classification: C91; D63; H23.

^{*}We would like to thank Jo Thori Lind, Andreas Kotsadam, Bertil Tungodden, Alexander Cappelen, Erik Sørensen, Matthias Sutter, Karine Nyborg, Kalle Moene, Steffen Huck, Henning Øien, Lea Heursen, Robert Stueber, Kai Barron, Gøril Andreassen, Dirk Engelmann, Kjell Arne Brekke, Vegard Wiborg, Tilman Fries, Thomas McKay and Siv-Elisabeth Skjelbred for valuable comments and discussions. 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 athletes deserve their equally high pay - observers insist. Accordingly, Lionel Messi and LeBron James encounter little criticism while top executives like Jamie Dimon and Larry Fink face skepticism and negative headlines when it comes to their high pay. Why do similarly high wages cause different reactions?

Research has shown that the source of inequality matters for legitimacy, with differences in pay based on merit and effort enjoying more legitimacy than differences based on luck (see e.g. Cappelen et al., 2007). The public might perceive that athletes high market value reflects merit and effort, while stories about CEOs self-proposed remuneration packages chip away at the legitimacy of their high pay.

Recent research suggests that there is substantial monopsony power in the labor market (Berger et al., 2022; Card, 2022). Thus, line managers assigning bonuses to factory workers or CEOs setting base wages for employees may exercise some power over wages. If they use their power for personal gain, their decisions can seem illegitimate. Inequality may signal potential power abuse in a setting with high discretion. Therefore, critical takes on inequality may reflect concerns about power abuse rather than concerns about inequality itself.

We study how the use and abuse of power matters for the legitimacy of inequality by conducting a large-scale experiment with 6000 participants on Prolific. Our experimental setting allows us to disentangle whether people object to power abuse or inequality. In observational settings, power abuse resulting in increased inequality makes it hard to discern these concerns. Moreover, power abuse may be hard to verify. For example, did the CEO earning high bonuses while cutting base salaries steer their company through a difficult crisis or did they merely persuade the board to adopt a favorable performance measure? Finally, individuals obtain power positions through processes that may themselves lend legitimacy, counteracting concerns about power abuse. In our experimental design we isolate the partial effect of power by randomizing power positions while keeping inequality levels orthogonal to decision making power.

We measure legitimacy perceptions through redistribution choices made by participants in the role of impartial spectators, finding that power abuse plays an independent and substantial role for inequality acceptance.

Following earlier studies on inequality and legitimacy perceptions, we run a two-stage experiment where workers produce in the first stage, and spectators redistribute in the second stage. In the production stage, randomly-paired workers solve a real effort task. We then divide worker pairs into two groups, matching worker pairs by performance across groups to obtain two identical performance distributions. In the first group, we randomly select one worker (henceforth "dictator") to decide initial earning shares for both workers in the pair. Then, we assign earnings determined by the first group's dictators to the performancematched worker pair in the second group. This ensures that both relative performance and initial earnings shares are identically distributed across the two groups.

In the redistribution stage, another set of participants, impartial spectators, learn about performance and initial earnings shares and set the final earnings. We assign spectators randomly to either the Dictator Decides Treatment (DD) or the External Decision Treatment (ED). In the DD treatment, spectators observe worker pairs where the dictator decided initial earnings. In the ED treatment, spectators observe worker pairs whose initial earnings are set by a dictator from a performance-matched pair, referred to as a third party in the instructions. After learning about initial earnings and performance, spectators in both treatments can redistribute, that is, they set final earnings. Crucially, the only difference between treatments is who decides earnings, and consequently whether spectators may witness and potentially object to power abuse.

Impartial spectators may desire to redistribute in this setting if they have social preferences, that is, they care about the outcomes of others. If they do have social preferences, real implementation provides an incentive for them to redistribute. Our settings speaks to two possible types of preferences, and our treatment isolates one of them. Firstly, spectators may be concerned about workers final outcomes and hold an ideal point about the desired final allocation. This may be, for example, meritocratic, egalitarian or some other pay. Secondly, spectators may have opinions about appropriate worker behaviour, and disapprove of choices where workers leverage their decision-making power to their own benefit. Spectators may be motivated by none, either one, or both motives and make a redistribution decision accordingly. Treating spectators with worker decision making power without varying inequality implies that any differences in spectator behaviour across treatment stem from concerns about power abuse.

This paper has three main findings. Firstly, spectators give lower final earnings to workers in a power position compared to workers whose initial earnings are decided externally. On average, spectators assign worker dictators in the Dictator Decides (DD) treatment 4 percentage points lower earnings than their counterpart in the External Decision (ED) treatment.

Secondly, the average effect hides strong interactions with the initial earnings choice. Spectators redistribute heavily when witnessing worker dictators that use their power to choose initial earnings shares exceeding their performance share.

Their counterparts in the ED treatment, facing equally overproportional initial earnings, but from a third party rather than a self-serving dictator worker, redistribute much less. In contrast, worker dictators choosing earnings shares equal to their performance (meritocratic earnings) receive the same final earnings as their counterparts. Spectators verbal reasoning indicates that they perceive dictator workers who claim overproportional earnings as selfish and unfair, while they do not mention any words connected to fairness in the ED treatment.

Thirdly, we investigate heterogeneity in spectators redistribution choices within treatment groups. When dictators exploit their power position by allocating an excessive earnings share to themselves, a third of DD spectators respond by assigning final earnings below their performance share, while the remaining DD spectators tend to correct to performance. In contrast, spectators in the ED treatment consistently correct excessive earnings by assigning the performance share. We interpret assigning earnings below performance as punishment for exploiting one's power position, given the otherwise strong drive towards meritocratic earnings in the experiment. Spectators switching from correcting to punishing behavior when faced with power abuse tend to self-classify as politically left-leaning and have high socio-economic status.

We contribute to the growing body of research on how the income process influences redistribution preferences (see for example Alesina and La Ferrara, 2005; Almås et al., 2020; Cappelen et al., 2007, 2022; Gärtner et al., 2017; Mollerstrom et al., 2015). A consistent finding in this literature is that people accept inequality more readily when inequality results from difference in effort or merit rather than luck. We go beyond the question on how different rule-based income schemes matter for redistribution preferences by letting individuals decision-making power matter for inequality. Opening for decision-making power allows us to study reactions to power use and abuse directly. Klimm (2019) and Bortolotti et al. (2017) allow for discretion in the specific form of cheating and find that the suspicion of cheating decreases inequality acceptance. By allowing use of power more generally, our setting transfers to a broader range of real-world wage setting processes. In a survey covering representative samples from 60 countries, Almås et al. (2022) document that people demand more redistribution in countries with widespread beliefs that inequality results from selfservingness. By measuring how spectators react to exogenous variation in the possibility of acting self-servingly, we can provide a causal test of the relationship between self-servingness and demand for redistribution. We find that spectators redistribute more when dictators use their power position for their own benefit at the expense of their co-worker.

Additionally, our work offers a new perspective on the recent renewed interest in wage-

setting power (for example Azar et al., 2022; Berger et al., 2022; Card, 2022). Acknowledging that firms may have more wage-setting power than previously imagined, what are the implications for accepted levels of inequality? The results from our experiment suggest that abuse of power significantly reduces inequality acceptance, indicating that (perceived) wage setting power may increase demand for redistribution.

In the next section, we detail the experimental design and give an overview of the data. In the result section, we first describe general features of spectator behavior before we present the main specification. We then move on to explore the mechanism and investigate heterogeneity in the difference in spectator choices between the two treatments before we conclude in the last section.

2 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, while in the spectator stage, another set of participants ("spectators") made distribution choices.¹ 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 (Almås et al., 2020; Bartling et al., 2018; Cappelen et al., 2019, 2022; Klimm, 2019).

The study was pre-registered in the AEA RCT-registry. We specified data collection and empirical estimation separately for each stage.² We ran both stages of the experiment online and recruited participants via Prolific.³ Participants learned that their answers would be

 $^{^{1}}$ We programmed the experiment in Lioness, a web-based platform for interactive online experiments (Arechar et al., 2018).

²Knutsen, Tora and Sonja Kovacevic. 2021. "To give or not to give. Explaining earnings choices in an online labor market experiment." AEA RCT Registry. October 15. https://doi.org/10.1257/rct.8345-1

³The platform differs in some important ways from the more commonly used Amazon MTurk: The platform encourages requesters 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

collected anonymously, that they could opt out of the study at any point in time and that we would not use deception.

2.1 Worker stage

We recruited 4000 participants for the worker stage, but due to 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).

$-109060000302405090060895489808002404137950204173090219495\\-109087144045705001024152700392166306020190800850199083960584177$

Figure 1: Example of number series we showed workers and asked them to count zeroes.

After completing the task, workers were informed about their own and the other workers relative performances. 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 their co-worker be able to review their choices.⁴ However, we did not specify to dictators what would happen if their choice was not selected for implementation.⁵ Workers received

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 et al., 2015)

⁴Researchers can reject participants submissions for completing surveys exceptionally fast or when 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.

⁵If they anticipated that there would be a third-party reviewing their choice, workers may 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' behaviour and not

payments within a few days after spectators had made decisions for the 50% of randomly selected cases where their choice was implemented.

Creating matched performance and initial choice groups

We construct two groups of worker pairs with identical performances and initial choices. Thus, spectators across the two treatments see worker pairs identical in performance and initial wage, only different in who determined the initial wage. In order to obtain this, 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 worker pair" in the second group.⁶

We equalize initial earnings by using choices made in the first group to set initial wages for performance-corresponding worker pairs in the second group. Thus, the dictator decides initial earnings both in their own pair *and* for a worker pair in the second 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.

worker choices by themselves.

⁶A 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.



Figure 2: Constructing performance-inequality matched worker samples

Workers performance follows a normal distribution (Figure 3a)⁷ 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 equally are the two most common choices (Figure 3b). Figure 3c shows that claiming earnings equal to one's performance is a third common pattern of behaviour. Figure 3d shows that average 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%. Defining dictators power abuse as claiming earnings exceeding one's performance share, 66% of choices fall in this category.

 $^{^{7}}$ We observe a slightly elevated fraction of 0/100% pairs, these are cases where one participant did not spend time on the tasks.





(c) Scatter plot workers' chosen shares and per- (d) Mean chosen share and share of workers formance choosing 50 and 100

Notes: All figures show raw data from the first stage of the experiment. The histogram in the upper right panel (a) shows all workers' shares of production. As this share is relative to the other worker the pair, the mean is 50%. The histogram in the upper right panel (b) shows the choices for the workers who got to choose their own earnings share. Panel (c) shows a scatter plot of workers' performances and choices of earnings shares. Panel (d) shows how i) the average choice, ii) the share choosing 50% and iii) the share choosing 100% vary with performance in terms of relative production share.

2.2 Spectator stage

We recruited 1962 spectators, one for each worker pair, between October 21 and 25, 2021. 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. Additionally, we invited some participants from less-studied regions to be able to study spectator responses across different institutional setting and inequality levels.⁸ A pilot study revealed that participants from some areas were much faster at entering our study than others. Therefore, we ran 7 studies simultaneously, each restricted to an area and with a fixed share of total participants per area: a) South Africa (8.5%), b) Mexico (8%), c) U.S (27%), d) U.K (10%), e) Western Europe (22%), f) Scandinavia (8.5%), and g) Eastern and Southern Europe (15%).⁹

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 Allocation (DD) treatment, spectators learn that a worker *within* the pair had decided their own earnings share. In the Assigned Allocation (ED) treatment, spectators learn that a *third party* had decided earnings shares for both workers.¹⁰ As is common in online experiments, we conducted comprehension checks to ensure subjects were attentive.¹¹ The instructions emphasized that spectators were making choices for real peo-

⁸There is a large literature on the sources of differences in redistribution preferences between (Western) Europe and the U.S (e.g. Alesina and Angeletos, 2005; Alesina and La Ferrara, 2005; Alesina et al., 2018) and also Scandinavia and the U.S. (e.g. Almås et al., 2020; Edlund, 1999; Kleven, 2014).

⁹In 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 increasing spots in the studies slowly.

¹⁰We do not tell spectators anything about the identity of this third party.

¹¹Subjects 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.

ple, adding information on how and when workers would be paid. Spectators then choose final earnings in both worker pairs.¹² Moving a slider allowed spectators to allocate relative earnings between the two workers. Spectators then proceed to reason 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.¹³

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. 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 may 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 decisions.

¹²Spectators choose for both worker pairs under the same treatment condition. As the two choices coming from the same spectator are not independent from each other, all regressions include standard errors clustered at the individual level.

 $^{^{13}\}mathrm{Appendix}$ A.1 includes the full set of screen shots from the experiment and the background questions are listed in Appendix A.3

	Dictator		Assi	Assigned		
Variable	Mean	sd	Mean	sd	p-value	
Matched varia	ables					
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 va						
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						
Right	2.05	0.90	2.06	0.87	0.98	
Ladder	5.49	1.60	5.50	1.65	0.86	
Time taken	654.02	322.51	615.92	287.77	0.01	
Ν	981		981			

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

2.3 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 Allocation (DD) treatment(b) Assigned Allocation (ED) treatmentFigure 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 dictator workers abuse their power to achieve earnings above their performance share.

3 Results

In this section, we first summarize general trends in spectator behaviour before moving to the main analysis of the treatment effect. Spectators assign on average 4 percentage points lower earnings in the DD treatment compared to the ED treatment. We demonstrate that spectators who face initial earnings choices exceeding relative performance drive this difference. Moreover, we observe heterogeneous responses to overproportional wages, with some spectators punishing dictators choosing wage shares above their performance share by assigning them shares *below* their performance share. In the final part of this section, we investigate the characteristics of those who punish and which dictator choices trigger punishment.

3.1 An overview: Spectator decision-making

Figure 5a 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 .

Figure 5: Mean spectator choice by worker performance



Notes: All figures show the smoothed mean function with 95% confidence intervals. The stippled line 45 degree line in the left panel serves as a reference point for wage shares equal to the performance share.

A large majority, 81%, of spectators propose final earnings different to the initial earnings they are presented, with 71% deviating more than 5 percentage points.¹⁴ 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.

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.

Figure 5b 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

 $^{^{14}}$ Spectators could choose the initial earnings on the slider but there was no default to keep initial earnings. Spectator used a slider that started randomly at 0% or 100% and had to move the slider to proceed

for meritocratic initial pay, the confidence interval is large for choices below 25%.

Spectators do not accept equal splits, referred to as "egalitarian choices", to the same extent as meritocratic earnings. These are cases were earnings are split 50%/50% regardless of performance. Overall, 46% of spectators in the ED treatment and 34% in the DD treatment did not make changes to egalitarian choices. Spectators reactions depend on whether the worker dictator (or their dictator counterpart) is the relatively worse or better performer with more acceptance of egalitarian choices for worse performers. In these cases, spectators award the low-performing worker a share slightly above their performance but still below 50%. For equal splits with the dictator's (or dictator counterpart) performance exceeding 50%, spectators differ in acceptance rates across the two treatment. Acceptance rates for DD and ED spectators are 33% and 55% respectively. In the ED treatment, spectators adjust the final wage share upwards towards the performance share, resulting in a pay scheme closer to a meritocratic allocation. In the DD treatment, spectators tend to accept higher performing dictators who suggest egalitarian pay. Answers in free text questions suggest that spectators in the DD treatment want to honour generous decisions of high-performing workers and consider the dictator worker entitled to share equally with their co-worker in spite of superior performance.¹⁵

In contrast to the relatively high acceptance rates of both meritocratic and egalitarian choices, spectators generally make changes when initial earnings exceed performance shares with only 10% leaving earnings unchanged and 16% making changes smaller than 5 percent-age points. Only 5% of spectators accept that 100% of earnings go to the dictator worker or their counterpart. Thus, spectators tend to accept claims for pay directly proportional to performance, and to a lesser extent, claims for equal splits. Earnings claims exceeding performance and extremely unequal claims are largely rejected.

¹⁵In answers in free text questions asking spectators to reason their choices, spectators mention that they perceived high-performer egalitarian choices as "generous" or "kind" and that they want to honour or respect those choices.

Having established general patterns of spectator behaviour, we turn to analyse differences in behaviour 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.

3.2 Regression Analysis

We start out by testing our main hypothesis; that mean final earnings in the DD treatment are lower than mean final earnings in the ED treatment.

In the following specification, we compare how much spectators give the dictator worker choosing their own earnings in the DD treatment, to how much spectators in the ED treatment give the corresponding worker with the same performance and initial earnings share. Spectators in the DD treatment may be concerned with the self-interest of workers choosing their own wage. Thus, we hypothesize lower mean final earnings in the DD treatment compared to mean final earnings in the ED treatment. 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 final earnings at the spectatorchoice level in %. 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, measuring the difference in final wage shares between DD and ED treatment. Based on our hypothesis we expect a negative γ , reflecting lower average wages in the DD group. **X** is our matrix of control variables, and includes time- and batch dummies¹⁶, as well as the performance share.

¹⁶We pre-registered to add dummies for batches of data collection, so that we could run additional rounds of data collection in case spectators drop out during the experiment. Running additional rounds meant that

DV: Final wage share assigned by spectator							
	(1)	(2)	(3)				
DD treatment	-3.99^{***}	-3.99^{***}	-3.99^{***}				
	(0.62)	(0.51)	(0.51)				
Performance		0.60^{***}	0.60^{***}				
		(0.02)	(0.02)				
Time Controls	No	No	Yes				
Batch Controls	No	No	Yes				
\mathbb{R}^2	0.01	0.36	0.36				
Adj. \mathbb{R}^2	0.01	0.36	0.36				
Num. obs.	3924	3924	3924				

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

***p < 0.01; **p < 0.05; *p < 0.1

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. All control variables were specified in our pre-plan. All standard errors are clustered at the individual level.

The first column in Table 2 reports the raw difference between the DD and ED treatment.¹⁷ We find that spectators assign 4 percentage points lower wages to dictator workers compared their counterpart. The results do not change controlling for performance (column 2) nor 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. Figure 6 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.

The average difference in final earnings between the DD and ED treatment results solely

we could collect a wage decision for all worker pairs in our sample.

¹⁷In Appendix Table B1a we show that the main results are robust to restricting the sample to only spectators' first decisions



Figure 6: Average effect by country or region

Notes: The plot shows the estimated γ coefficient separately by country/group of countries, i.e. the difference in assigned wages between the DD and ED treatment from Equation (1). All regressions include a linear term controlling for performance and standard errors clustered at the individual level.

from who is deciding wages, as both initial choices and performances were identically distributed across the two groups.¹⁸ However, the average effect masks heterogeneous responses to different initial earnings as well as heterogeneity in spectators' reactions given the initial choice.

Next, we investigate which types of worker choices drive the treatment. For this purpose, we classify initial choices as overproportional, meritocratic or underproportional. We refer to initial allocations proportional to performance as meritocratic, allocations exceeding performance as overproportional and allocations below performance as underproportional.

In the regression below, we measure the concern about self-servingness as the interaction effect between treatment status and overproportional choices. Similarly, the interaction effect between treatment status and underproportional choices picks up spectators assessment of generous choices. The regression below includes the interactions between the DD treatment dummy 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).

$$w_{c} = \alpha + \beta X_{ci} + \gamma T_{i} + \delta T_{i} * Overprop_{c} + \theta T_{i} * Underprop_{c} + \pi Overprop_{c} + \nu Underprop_{c} + u_{i}$$

$$(2)$$

Overproportional is the indicator variable for overproportional choices, while Underproportional is the indicator variable for underproportional choices. δ and θ are the coefficient of interests for the interactions between treatment group and the proportionality indicators. The

¹⁸We measure fairness perceptions of our spectators as their willingness to redistribute earnings between the workers vis--vis the initial allocation. In order to pick up fairness perceptions in this way, spectators need to be willing to intervene. Ambuehl et al. (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 behaviour. 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.

DV: Final wage share assigned by spectator							
	(1)	(2)	(3)				
DD treatment	-2.48^{***}	-0.04	0.35				
	(0.57)	(2.02)	(1.11)				
Performance	0.57^{***}	0.64***	0.59***				
	(0.02)	(0.03)	(0.02)				
Take-rate	0.00	· · · ·	· · · ·				
	(0.01)						
Overproportional			1.00				
			(0.82)				
Underproportional			-0.61				
			(1.00)				
DD: Take-rate	-0.08^{***}						
	(0.02)						
DD: Performance		-0.07^{**}					
		(0.04)					
DD: overprop			-5.75^{***}				
			(1.30)				
DD: underprop			-2.76^{*}				
			(1.49)				
\mathbb{R}^2	0.37	0.36	0.36				
Adj. \mathbb{R}^2	0.36	0.36	0.36				
Num. obs.	3924	3924	3924				

Table 3: Mechanism: How the treatment effect varies with performance and overproportionality

***p < 0.01; **p < 0.05; *p < 0.1

In this table all columns presents coefficients from regressions with spectators' assigned wage shares as the dependent variable. In column (1) *take-rate* refers to the to the difference between initial wage share and performance. In column (3) *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. In addition to control for performance, all regressions include a set of time of entry and batch dummies. Standard errors are clustered at the individual level. third column in Table 3 shows that as expected γ , that is the interaction between the DD treatment and over-proportional choice, is negative. 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 a third of a standard deviation.

The second column in the same table shows that the treatment effect also increases with the take-rate, i.e. the difference between performance and initial wage, as the interaction term between take-rate and DD treatment is negative and significant at the 1% level. Thus, the more overproportional initial wages are relative to the performance share, the larger is the difference between final wages in the DD and ED treatment. Interestingly, the coefficient on take-rate for the external decision (ED) treatment is 0. Overproportional compared to proportional initial wages do not lead to lower final wages when initial wages are decided externally.

Figure 7 shows the words used relatively more frequent by the two treatment conditions when asked to verbally reason their choices. 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 dictator or "teach them a lesson". Spectators in the external decision treatment use neutral vocabulary and prominently mention "time", "productive", "effort" and "deserve". Appendix Table B2 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 assigned allocations compared to 58 times by spectators seeing worker dictators choices.

The difference in assigned wage shares between the two treatments is smaller for generous choices than for overproportional ones. Workers who choose lower-than-performance-pay for Figure 7: Wordcloud open-text question: Words used relatively more often by one treatment condition compared to the other



Notes: After deciding the final allocations, spectators had to reason their choices in an open text question. The figure shows a word cloud of words used relatively more frequent in one treatment group compared to the other. In Appendix B.2 we provide a frequency table as well as some details on the analysis of the open text question.

themselves do not receive significantly different wages compared to workers whose earnings were determined by a third party. If anything, workers receive lower final wages when offering a generous wage share to their co-worker. Spectators seem think that it is workers good right to give away some of their money (as indicated by their performance share), while the same choice seems unfair when coming from a third party.

Going back to overproportional initial choices, spectators in the ED treatment do not assign workers lower pay in these cases compared to cases with meritocratic shares. When the third party proposes overproportional wages, spectators decide on final earnings close to the average of all final earnings in the sample (51.25% vs. 51.31%), while workers proposing a similar split for themselves receive 45%. This does not mean that spectators in the ED treatment accept externally assigned overproportional claims, but that they *correct* to performance share. Spectators in the DD treatment depart from the meritocratic wage norm and assign wages on average below performance, most likely to reprimand choices that they consider as illegitimate.

Spectators in the DD treatment react particularly to substantially overproportional worker choices. As Figure 8 indicates, choices exceeding performance shares by up to 25 percentage points do provoke much of a differential reaction, while choices in the range from 25 to 75 percentage points differ substantially depending on who allocated initial earnings. The 45-degree stippled line in Figure 8 shows how workers receive a wage share close to their performance share when a third party proposes overproportional pay, while workers choosing their own pay face final earnings below their performance share.



Figure 8: 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. This figure is based on the sample of overproportional initial choices, i.e. cases where dictator worker or their shadow dictator counterpart had earnings shares exceeding performance shares. The stippled 45 degree line serves as a reference point for wage shares equal to the performance share.

3.3 Prevalence and determinants of punishment

In this section we focus on cases where spectators face overproportional initial choices. While spectators in the ED treatment seem to correct overproportional shares back the performance share, a sizable fraction of spectators in the DD treatment (35%) decides to pay dictators a final wage below their performance share. In fact, the average worker asking for overproportional pay receives substantially less than their performance share. We classify these spectator decisions as punishing. Here, we explore the frequency, intensity and drivers of punishment.

3.3.1 Prevalence

On average, spectators in the DD treatment assign lower wages when workers claim overproportional pay. We interpret this penalty as punishment for power abuse. Spectators may object to these choices since claiming an overproportional share for oneself mechanically translates in assigning one co-worker underproportional earnings.

We will 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".

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. Moreover, Figure B1 in the Appendix shows that countries with a significant treatment effect also had the largest difference in the share of punishing choices between the DD and ED treatment, while the countries without a treatment effect had no significant change in this share.

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 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 people 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 B3 that also 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. In the ED treatment, this relationship is reversed with punishment being negatively correlated with the mention of "greed" and positively with the use of "performance".

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, 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.

Accepting



Correcting

Figure 9: Wordcloud open-text question: Comparison of "punishers", "correcters" and "accepters" in the DD treatment

Notes: In an open-text question spectators had to reason their choices. The wordcloud shows words relatively more frequently used by spectators who made i) punishing,ii) accepting and iii) correcting choices in the DD treatment. These categories are defined according to spectators choices. "Punishers" are spectators whose choices can be defined by: $final \ wage < performance < initial \ wage$. "Correcters" make choices where $final \ wage = performance < initial \ wage$ and "accepters" choose a final wage greater than performance share. Note that the sample of spectators is restricted to those in the DD treatment who faced overproportional initial choices.





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.

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

3.3.2 Determinants

Citizens voting on labor market regulation and redistribution may face similar choices and trade-offs as the spectators in our experiment. Characterizing spectators willing to punish self-serving choices, we find that these spectators tend to be politically left-leaning and believe that justice will prevail over injustice eventually. Table 4 shows how the interaction between treatments and i) being left-winged, ii) believing that justice will prevail and iii) self-positioning on a ladder, predicts punishing and correcting behavior. The results indicate that in the DD treatment those punishing are more left-winged, and those correcting more right-winged than in the ED treatment. Note that the two treatment conditions are balanced with respect to political views as shown in Table 1. Therefore, these results suggest that those affected by our treatment, by moving from correcting to punishing behavior, tend to be more left-winged and believe that justice will prevail over injustice.

The variation in the inclination to make punishing choices also affects the overall treatment effect. Being left-winged and believing in justice are correlated with larger treatment effects, while background variables such as gender and age seem to be unrelated. Appendix Table B4 shows that interacting the treatment-dummy with being left-winged and believing that justice will prevail halves the 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" seems to be 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".

	(1)	(2)	(3)	(4)	(5)	(6)
	Punishment1	Correcting1	Punishment2	Correcting2	Punishment3	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)
leftRight	0.02	-0.02				
	(0.01)	(0.02)				
DD:leftRight	-0.05***	0.04^{*}				
	(0.02)	(0.02)				
justiceoverinjustice			0.00	0.03^{**}		
			(0.01)	(0.01)		
DD:justiceoverinjustice			0.03^{*}	-0.05***		
			(0.02)	(0.02)		
ladder					0.00	0.00
					(0.01)	(0.01)
DD:ladder					0.02**	-0.00
					(0.01)	(0.01)
\mathbb{R}^2	0.11	0.01	0.11	0.01	0.11	0.01
Adj. \mathbb{R}^2	0.10	0.01	0.10	0.01	0.10	0.01
Num. obs.	2565	2565	2588	2588	2587	2587

Table 4: Table showing how the interaction between DD treatment and answers to survey questions predict punishing or correcting behaviour

***p < 0.01; **p < 0.05; *p < 0.1 Column 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* and 0 otherwise, as the dependent variable. All regressions include only spectators facing overproportional initial choices and control for performance and initial choice.

4 Conclusion

In this paper, we present the results of a large-scale online experiment designed to explore how people react to income inequality arising from power abuse. The experiment comprised of two stages: First, we recruited one set of participants, workers, to perform a task and sorted them into pairs. Each pair received a bonus to be shared between them. We randomized whether one worker in the pair, the dictator, or an external third party made the decision on how to split the bonus, which we refer to as initial earnings. We designed the experiment so that the distribution of earnings in these two groups was identical. Next, we recruited another set of participants, impartial spectators, to make a final decision on earnings. We randomly assigned spectators to receive one out of two treatments. Spectators redistributed initial earnings of workers choosing their own initial earnings (DD treatment), or of workers whose earnings were decided externally (ED treatment).

Our findings show that spectators allocate less to dictator workers than to workers with identical, but externally decided earnings. Across the different treatments, most spectators choose earnings shares equal to performance. The difference between the two treatments stems from a minority of spectators assigning earnings below this meritocratic benchmark to dictators who have chosen initial earnings exceeding their performance share, compared to receiving equally overproportional earnings. Analyzing open-text answers, we find that these spectators more frequently than others mention words like "greed" and "selfishness" when reasoning their choice. Therefore, we interpret these choices as spectators punishing workers who abuse their power to pursue higher earnings at the expense of their co-worker.

We contribute to the literature on redistribution preferences by showing that power abuse in the income process affects inequality acceptance. In our experiment, acceptance for meritocratic pay decreases when we introduce the possibility to extract earnings at the expense of someone else. Although we acknowledge that we have removed many important features of inequality in society in our experiment, we believe that our findings contribute to the understanding 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 may have taken place. An interesting venue for future research would be to explore whether spectators also redistribute in a setting with uncertainty about power abuse.

References

- Ackfeld, Viola and Axel Ockenfels (2021). "Do people intervene to make others behave prosocially?" In: *Games and Economic Behavior* 128, pp. 58–72.
- Alesina, Alberto and George-Marios Angeletos (2005). "Fairness and redistribution." In: American economic review 95.4, pp. 960–980.
- Alesina, Alberto and Eliana La Ferrara (2005). "Preferences for redistribution in the land of opportunities." In: Journal of Public Economics 89.5-6, pp. 897–931.
- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso (2018). "Intergenerational mobility and preferences for redistribution." In: *American Economic Review* 108.2, pp. 521–54.
- Almås, Ingvild, Alexander W Cappelen, and Bertil Tungodden (2020). "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?" In: Journal of Political Economy 128.5, pp. 1753–1788.
- Almås, Ingvild, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden (2022).
 "Global evidence on the selfish rich inequality hypothesis." In: *Proceedings of the National Academy of Sciences* 119.3.
- Ambuehl, Sandro, B Douglas Bernheim, and Axel Ockenfels (2021). "What motivates paternalism? An experimental study." In: American economic review 111.3, pp. 787–830.
- Arechar, Antonio A, Simon Gächter, and Lucas Molleman (2018). "Conducting interactive experiments online." In: *Experimental Economics* 21.1, pp. 99–131.
- Azar, José, Ioana Marinescu, and Marshall Steinbaum (2022). "Labor market concentration." In: Journal of Human Resources 57.S, S167–S199.
- Bartling, Björn, Alexander W Cappelen, Mathias Ekström, Erik Sorensen, and Bertil Tungodden (2018). "Fairness in Winner-Take-All Markets." In: NHH Dept. of Economics Discussion Paper 08.

- Berger, David, Kyle Herkenhoff, and Simon Mongey (2022). "Labor market power." In: American Economic Review 112.4, pp. 1147–93.
- Bortolotti, Stefania, Ivan Soraperra, Matthias Sutter, and Claudia Zoller (2017). "Too Lucky to be True-Fairness Views under the Shadow of Cheating." In: *IZA Discussion Paper*.
- Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden (2007).
 "The pluralism of fairness ideals: An experimental approach." In: American Economic Review 97.3, pp. 818–827.
- Cappelen, Alexander W, Ranveig Falch, and Bertil Tungodden (2019). "The boy crisis: Experimental evidence on the acceptance of males falling behind." In: NHH Dept. of Economics Discussion Paper 06.
- Cappelen, Alexander W, Karl Ove Moene, Siv-Elisabeth Skjelbred, and Bertil Tungodden (2022). "The Merit Primacy Effect*." In: The Economic Journal. ueac082. ISSN: 0013-0133. DOI: 10.1093/ej/ueac082. eprint: https://academic.oup.com/ej/advancearticle-pdf/doi/10.1093/ej/ueac082/47884361/ueac082.pdf. URL: https: //doi.org/10.1093/ej/ueac082.
- Card, David (2022). "Who set your wage?" In: American Economic Review 112.4, pp. 1075–90.
- Charité, Jimmy, Raymond Fisman, Ilyana Kuziemko, and Kewei Zhang (2022). "Reference points and redistributive preferences: Experimental evidence." In: *Journal of Public Economics* 216, p. 104761.
- Edlund, Jonas (1999). "Trust in government and welfare regimes: Attitudes to redistribution and financial cheating in the USA and Norway." In: *European Journal of Political Research* 35.3, pp. 341–370.
- Gärtner, Manja, Johanna Mollerstrom, and David Seim (2017). "Individual risk preferences and the demand for redistribution." In: *Journal of Public Economics* 153, pp. 49–55.

- Kleven, Henrik Jacobsen (2014). "How can Scandinavians tax so much?" In: Journal of Economic Perspectives 28.4, pp. 77–98.
- Klimm, Felix (2019). "Suspicious Success: Cheating, Inequality Acceptance, and Political Preferences." In: European Economic Review.
- 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." In: *Behavior research methods* 47.2, pp. 519–528.
- Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø Sørensen (2015). "Luck, choice and responsibilityAn experimental study of fairness views." In: Journal of Public Economics 131, pp. 33–40.
- Palan, Stefan and Christian Schitter (2018). "Prolific.ac-A subject pool for online experiments." In: Journal of Behavioral and Experimental Finance 17, pp. 22–27.

Appendices

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



(c) Information about the task workers did

Figure A11: Information to spectators (same for all)



Figure A12: Information about initial earnings





(a) DD treatment

(b) ED treatment

Figure A15: Open-text question

You just made two decisions on how to split a bonus in a pair of participants.

In the first pair, participant A got 60% of the points while participant B got 40% of the points. Participant A decided to give 70% of the total pay-off to thermselves and 30% to the other participant. You decided for a final share of 50% to participant A.

In the second pair, participant A got 51% of the points while participant B got 49% of the points. Participant A decided to give 39% of the total pay-off to themselves and 61% to the other participant. You decided for a final share of 55% to participant A.

Please explain the reasons for your choices in the text field below. What was the guiding principle for your decisions?

You just made a decision on two decisions on how to split a bonus in a pair of participants.

In the first pair, participant A got 60% of the points while participant B got 40% of the points. A third-party participant decided to give 70% of total earnings to participant A, while you decided to give 60% to this participant.

In the second pair, participant B got 49% of the points while participant A got 51% of the points. A third-party participant decided to give 100% of the total pay-off to participant. B and 0% to the other participant. You decided for a final share of 50% to participant B.

Please explain the reasons for your choices in the text field below. What was the guiding principle for your decisions?

Continue

(a) DD treatment

(b) ED treatment

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:

Politician Football player CEO of a large company Taxi driver 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 theres 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 offthose 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 also if we look only at spectators' first decision. These results are shown Table B1a. Moreover, 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:

$$w_c = \alpha + \gamma T_i + \delta LowPerformer_c * T_i + \theta LowPerformer_c + \beta X_{ci} + u_i$$
(B1)

Where *lowPerformer* is an indicator equal to 1 if the preliminary earnings belong to the 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 B1b.

	(1)	(2)	(b) Table showing that the treatment effect				
DD treatment	-3.34^{***}	-0.53	is smaller when low performers choose wages				
	(0.63)	(1.61)		(1)	(2)		
Performance	0.60^{***}	0.61^{***}	DD treatment	-5.30***	-5.28***		
	(0.02)	(0.02)		(0.80)	(0.71)		
Overprop		1.01	Low performer	-18.56^{***}	-3.64***		
		(1.28)	-	(0.70)	(0.75)		
Underprop		-1.55	DD treatment: Low performer	2.86^{***}	2.88^{***}		
DD transformet, Original		(1.54)		(1.10)	(1.03)		
DD treatment: Overproportional		-3.79°	Performance		0.55^{***}		
DD treatment: Underproportional		(1.78) -1.56			(0.03)		
bb steament. Enderproportional		(2.17)	R^2	0.23	0.36		
	0.40	0.40	Adj. \mathbb{R}^2	0.22	0.36		
R^2	0.40	0.40	Num. obs.	3924	3924		
Adj. K ⁻	0.39	0.39	$***n < 0.01 \cdot **n < 0.05 \cdot *n < 0$	0.1 This ta	ble shows		
Num. obs.	1902	1902	p < 0.01, p < 0.00, p <	J.I. IIII5 0a	SHOWS		

(a) Main results using only spectators' first decisions

***p < 0.01; **p < 0.05; *p < 0.1. This table presents coefficients from regression 1 and 2 using only spectators' first decision. *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. Both regressions also include batch and time controls. the coefficients from estimating Regression B1. All regressions include batch and time controls as well as standard errors clustered at the individual level. lowPerformer is an indicator equal to 1 if the preliminary earnings belong to the the lower performer in the worker pair.

Table B1: Additional pre-registered regressions

B.2 Analysis open-text question

After spectators had made decisions on how to split earnings within the worker pairs, spectators had to reason their choices in an open-text question. The screenshots in Figure A15 show how the spectators were asked to reason 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. R package "stopwords" provide a list of such words. This is 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 to remove some other words that we saw people used in the pilot, but that gave little information.¹⁹

Next, we make a frequency table by treatment group, this means that for each word we count how many times it was used in the DD and ED treatment. We keep only those words that are mentioned 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 B2 shows the top 20 words relatively more frequently used by spectators in one treatment group compared to the other.

¹⁹These 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".

	ED	DD	Relative freq.		ED	DD	Relative Freq
greed	0	16	Inf	thirdparty	23	0	Inf
selfish	0	60	Inf	round	19	2	9.50
honest	2	11	5.50	party	23	5	4.60
willing	2	11	5.50	third	22	6	3.67
instead	4	19	4.75	difference	81	23	3.52
choose	6	26	4.33	productive	15	5	3.00
partner	4	17	4.25	rewards	15	5	3.00
initially	3	12	4.00	compensated	17	6	2.83
splitting	5	19	3.80	attention	11	4	2.75
payoff	36	130	3.61	earnings	72	28	2.57
decide	6	20	3.33	harder	12	5	2.40
winnings	4	13	3.25	compensation	20	9	2.22
though	18	57	3.17	big	19	9	2.11
nothing	8	25	3.12	numbers	18	9	2.00
take	16	46	2.88	outcome	15	8	1.88
matter	6	17	2.83	survey	11	6	1.83
agreed	5	14	2.80	effort	102	58	1.76
agree	4	11	2.75	different	19	11	1.73
keep	10	27	2.70	mean	12	7	1.71
unfair	20	54	2.70	accordingly	13	8	1.62

Table B2: Top 20 words used relatively more frequent in one treatment group compared to the other

(a) More frequent DD treatment

(b) More frequent ED treatment

Notes: This table shows the frequencies of words used relatively more often in a) the DD treatment compared to the ED treatment and b) the ED treatment compared to the DD treatment. All words are extracted from an open-text question where spectators reasoned their distribution decision. We ranked the words by relative frequency, i.e. how many times they were used in one group compared to the other, and these tables show the lists the top 20 words relatively more frequently used by one treatment compared to the other.

DV: Punishing	(1)	(2)	(3)
	All	Within DD	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)
Mention greed	-0.11^{***}	0.29^{***}	-0.18^{***}
	(0.04)	(0.04)	(0.03)
Mention performance	0.09^{***}	-0.06^{*}	0.09^{***}
	(0.03)	(0.03)	(0.03)
Mention equal	0.13^{***}	0.07^{**}	0.13^{***}
	(0.04)	(0.04)	(0.04)
DD: mention greed	0.41^{***}		
	(0.05)		
DD: mention performance	-0.15^{***}		
	(0.05)		
DD: mention equal	-0.07		
	(0.05)		
\mathbb{R}^2	0.14	0.19	0.07
$\operatorname{Adj.} \mathbb{R}^2$	0.14	0.19	0.07
Num. obs.	2610	1305	1305

Table B3: Table showing how the use of "greed/selfishness", "performance" and "equal" predict punishing choices

Notes: ***p < 0.01; **p < 0.05; *p < 0.1 In this table mention greed refers to a dummy variable equal to 1 if a spectator mentions the words "greed", "greedy", "selfish" or "selfishness" when reasoning their choices, and 0 otherwise. Similarly, mention performance is a dummy equal to one if spectators mention "performance" or "score" and mention equal is a dummy equal to 1 if spectators mention "equal" or "equally". Note that the sample includes only spectators facing overproportional initial choices. Standard errors are clustered at the individual level.

B.3 Additional results

Figure B1: Change in share of punishing choices between ED and DD treatment by country/region



Notes: The figure shows the change in the share of punishing choices between the DD and ED treatment by country/region level. Punishing is defined as *final wage < performance < initial wage*. Note that the share of punishing choices is calculated only for spectators facing overproportional initial choices (i.e. *performance < initial wage*.

	(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	-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:left	-1.80					
	(1.16)					
Female		0.23				
		(0.64)				
DD:female		1.13				
		(1.05)				
Age			-0.03			
			(0.03)			
DD :age			0.00			
			(0.03)			
Inequality				0.02		
				(0.11)		
DD:inequality				0.12		
				(0.19)		
tryfair					-0.16	
					(0.13)	
DD:tryfair					0.47^{**}	
					(0.21)	
justiceoverinjustice						-0.48*
						(0.29)
DD:justiceoverinjustice						-0.58
						(0.49)
Pr(>F)	0.033	0.13	0.98	0.46	0.014	0.22
\mathbb{R}^2	0.36	0.36	0.36	0.36	0.36	0.36
Adj. \mathbb{R}^2	0.36	0.36	0.36	0.36	0.36	0.36
Num. obs.	3856	3894	3880	3898	3888	3886

Table B4: Table showing the treatment effect interacted with background variables and answers to survey questions

Notes: ***p < 0.01; **p < 0.05; *p < 0.1 This table shows the results from a set of regressions where the treatment-dummy is interacted with background variables and answers to survey questions. In column (1) Left-winged refers to a dummy variable equal to 1 for individuals who in our survey respond being "left"/"center-left" and 0 otherwise. In column (2) Female refers to a dummy equal to 1 for females and 0 for males and in column (3) Age is a continuous variable for age. In the last three columns the treatment-dummy is interacted with answers to survey questions on whether incomes should be made more equal (*inequality*) on a scale from 1 to 10, whether most people would take advantage of you (1) or try to be fair (10) (tryfair) and whether justice will prevail over injustice (Justice) on a scale of 5 from "strongly agree" to "strongly disagree". The full questions are listed in Appendix A.3. The Pr(>F) refers to the probability from a F-test comparing the50 odel with the interaction term to the model without.In addition to controlling for performance, all regressions also include time and batch controls.Standard errors are clustered at the individual level.