Discussion Paper Series – CRC TR 224

Discussion Paper No. 318
Project A 01

Shallow Meritocracy

Peter Andre [1]

October 2022
(*First version: September 2021,*
Second version: February 2022)

[1] University of Bonn, Email: p.andre@uni-bonn.de

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Shallow Meritocracy

Peter Andre

September 28, 2022

Abstract: Meritocracies aspire to reward hard work but promise not to judge individuals by the circumstances into which they were born. However, circumstances often shape the choice to work hard. I show that people's merit judgments are insensitive to this effect. They hold others responsible for their choices, even if these choices have been shaped by unequal circumstances. In an experiment, US participants judge how much money workers deserve for the effort they exert. Unequal circumstances disadvantage some workers and discourage them from working hard. Nonetheless, participants reward the effort of disadvantaged and advantaged workers identically, regardless of the circumstances under which choices are made. For some participants, this reflects their fundamental view on fair rewards. For others, the neglect results from the uncertain counterfactual. They understand that unequal circumstances shape choices but do not correct for this because the exact counterfactual—what would have happened under equal circumstances—remains uncertain.

JEL-Codes: C91, D63, D91, H23.

Keywords: Meritocracy, fairness, responsibility, attitudes towards inequality, redistribution, social preferences, inference, uncertainty, counterfactual thinking.

Date of first version: September 4, 2021

Contact: Peter Andre, briq – Institute on Behavior & Inequality (Bonn, Germany), peter.andre@briq-institute.org. **Acknowledgements**: I thank Ingvild Almås, Teodora Boneva, Alexander Cappelen, Felix Chopra, Thomas Dohmen, Armin Falk, Arkadev Ghosh, Thomas Graeber, Ingar Haaland, Leander Heldring, Luca Henkel, Samuel Hirshman, Paul Hufe, Claus Kreiner, Yucheng Liang, Matt Lowe, Wladislaw Mill, Maximilian Müller, Suanna Oh, Franz Ostrizek, Christopher Roth, Sebastian Schaube, Erik Sørensen, Andreas Stegmann, Bertil Tungodden, Johannes Wohlfart, Florian Zimmermann, and participants at various conferences and seminars for helpful comments and discussions. **Funding**: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. Funding by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged. Supported by the Reinhard-Selten Scholarship (German Association for Experimental Economic Research). Supported by the Joachim Herz Foundation. **Ethics approval**: The study obtained ethics approval from the German Association for Experimental Economic Research (#HyegJqzx, 12/11/2019). **Research transparency**: The study was preregistered at the AEA RCT Registry (#AEARCTR-0005811). Data and code will be made available. I declare no competing interests. See also Appendix F on research transparency. **Instructions**: The experimental instructions of all studies are available at https://osf.io/xj7vc/.

1 Introduction

The notion of meritocratic fairness is at the heart of Western political and economic culture. It shapes which inequalities are considered fair, which redistributive policies are implemented, and how welfare states are designed (Alesina and Glaeser, 2004; Alesina and Angeletos, 2005; Cappelen et al., 2020b; Sandel, 2020). In essence, meritocratic fairness means that people should be rewarded in proportion to their merit, and—besides talent and skill—the choice to work hard and exert effort is considered particularly meritorious. On the contrary, external circumstances such as parental background, race, or sex are not viewed as legitimate sources of merit (Almås et al., 2020; Konow, 2000; Roemer, 1993). Meritocratic fairness thus distinguishes between effort choices (relevant for reward) and external circumstances (irrelevant for reward).

However, this fundamental distinction between choices and circumstances is clouded by a ubiquitous feature of human behavior: Agents' choices are shaped by the circumstances, opportunities, and incentives they face. For example, a person growing up with few opportunities and incentives to work hard might respond by exerting little effort. Likewise, minorities who experience discrimination may be discouraged from working hard. In fact, empirical studies have linked effort, career, and schooling choices with socioeconomic inequalities (e.g., Altmejd et al., 2021; Bursztyn et al., 2017; Carlana et al., 2022; Falk et al., 2020a; Glover et al., 2017; Müller, 2021; Sviatschi, 2022). And the fact that adverse environments often lead to detrimental decision-making is considered an important cause of poverty (e.g., Bertrand et al., 2004; Haushofer and Fehr, 2014).

Thus, a fundamental issue in any meritocracy is how to reward choices that are shaped by unequal external circumstances. Are people held responsible for choices that result from unequal circumstances? This study explores the prevailing notion of meritocratic fairness in the United States and investigates whether people reward choices in the light of or irrespective of the surrounding circumstances.

Answering this question requires the tight control of an experimental set-up. The ideal test compares how people reward choices made under different circumstances if, *ceteris paribus*, the different circumstances did or did not influence which choices were made. I create this variation in a series of allocation experiments with a large, broadly representative US sample of more than 6,000 respondents. The study proceeds in three steps. First, I show that people's merit judgments do *not* factor in the circumstances under which choices are made, even in a simple and transparent allocation setting. Second, I explore the behavioral mechanism, exploiting the precise control of the experimental setting. Finally, I confirm these patterns in real-world examples of unequal circumstances with the help of a complementary vignette study.

In the main experiment, each participant ("spectator") judges how much money two "workers" should earn for their effort in a piece-rate job. Workers work on a standardized task, and their *effort choice* is how many tasks they complete. Their *circumstances* are their exogenously determined returns to effort, that is, the piece rate they earn that is randomly assigned and can be either high ($0.50) or low ($0.10), each with a 50% chance. By chance, one worker receives the high rate, and the other the low rate. All workers know about the lottery, but—as described below—I vary across treatments whether workers know their assigned rate. The spectators are fully informed about the workers' situation, then they decide what final payment each worker should earn. They can freely redistribute the earnings between the two workers, thus judging which reward each worker deserves. These merit judgments are the central outcome variable of the study. Spectators make multiple merit judgments under different scenarios, each presenting different effort choices that workers could make. To incentivize spectators' decisions, a random subset of their redistribution decisions is implemented.

To identify whether spectators' merit judgments take into account that workers' effort choices are shaped by their random circumstances, the experiment exogenously varies the environment in which workers make their effort choices. In the *control* condition, the workers do not yet know their realized piece rates. They only know their odds of obtaining a high or low piece rate, which are identical for both workers. Hence, their effort choices are directly comparable because their choices are made in the same environment—a level playing field. By contrast, in the *treatment* condition, workers immediately learn about their realized piece rates. Workers with a high piece rate are encouraged to work hard, whereas workers with a low piece rate are discouraged to work hard. Indeed, workers complete roughly three times as many tasks for the high than for the low piece rate. Thus, circumstances differentially shape workers' choices in the treatment condition but not in the control condition.

I compare the merit judgments of spectators across the two conditions and test whether spectators compensate disadvantaged workers in the treatment condition for the fact that they are discouraged from working hard. The results show that the merit judgments of participants do not factor in that unequal circumstances shape the choices of workers. While spectators redistribute payments to reward workers for greater effort, they do so equally in both conditions. Thus, the disadvantaged worker is not compensated for facing discouraging circumstances. A large sample size allows me to rule out even minor increases in the reward of the disadvantaged worker (0.8 percentage points of total payoff). The results thus provide strong evidence for the absence of a meaningful effect. Spectators hold workers responsible for their choices, even if these choices are shaped by external circumstances over which the workers have no control.

Next, I ask *why* spectators do not factor in that circumstances influence the choices of workers. I start by investigating whether spectators underestimate the effect of circumstances on effort choices, in line with the fundamental attribution error (Ross, 1977). If spectators underestimate the effect, they have little reason to correct for it. I measure incentivized beliefs about how strongly the piece rates influence the effort choices of workers. Inconsistent with a fundamental attribution error, the results show that spectators even slightly overestimate the piece-rate effect. Of course, this does not imply that spectators also pay sufficient attention to it while rewarding the workers. In an additional experimental condition, I therefore implement an attention intervention in which I draw spectators' attention to the effect of circumstances just before their merit judgments. However, even then, their merit judgments remain insensitive to the effect of circumstances on choices.

Thus, spectators seem to be aware of and accurately anticipate the average expected piece-rate effect. However, they still do not know with certainty what the two specific workers for whom they are responsible would have done in equally advantaged circumstances. Would their disadvantaged worker have worked much harder for the high piece rate, or would he still have exerted only little effort? This specific counterfactual remains unknown and uncertain, even when the expected counterfactual is known. I show that, in light of this uncertainty, spectators base their merit judgments on what they know with certainty: observed effort levels. For this purpose, I conduct an additional experiment in which I exogenously resolve the uncertainty of the counterfactual. I provide a subset of spectators with accurate and reliable information about what their specific disadvantaged worker would have done in the advantaged environment. I find that the average merit judgments of spectators react strongly to this information. Once spectators have "hard" evidence that their disadvantaged worker would have worked more in the advantaged environment, they take the effect of circumstances into account and compensate the disadvantaged worker. By contrast, spectators who do not receive any information about the counterfactual remain unresponsive.

However, crucially, this effect appears to be driven by a subset of spectators. The merit judgments of spectators are very heterogeneous even when the counterfactual state is known, and they often align with distinct fairness views. This is no coincidence. When asked to describe the rationale behind their reward decision, spectators explicitly refer to different fairness notions. In particular, I distinguish between two different meritocratic fairness views to which many respondents refer: "comparable choice meritocratism" and "actual choice meritocratism". Comparable choice meritocrats think that merit judgments should be based on the counterfactual effort choices that workers would make in identical comparable circumstances. Actual choice meritocrats think workers should be rewarded proportionally to their actual effort choices, even if these choices are shaped by unequal circumstances. To assess the prevalence of these different merit views in the population, I estimate a simple structural model. The model classifies 28% of participants as comparable choice meritocrats and 40% of participants as actual choice meritocrats. In addition, I estimate a share of 16% "libertarians" who accept any inequality and never redistribute and 15% "egalitarians" who think that the workers always deserve equal payment. The results show that people hold fundamentally different fairness views. Importantly, they also reveal that, even in the rare case where counterfactual choices are known, only about one-quarter of individuals would factor in that unequal circumstances shape the choices of agents.

Although the controlled experimental environment has the crucial advantage that the effect of interest is credibly identified, it also comes at a cost. It differs from many real-life settings that characterize the debate about merit, choices, and circumstances. In the third and final step, I therefore run a vignette study and show that the insensitivity of merit judgments to the effect of circumstances on choices can also be observed in relevant labor market and career choice scenarios. For example, participants do not compensate a black employee who chooses not to work hard for a promotion but faces racial discrimination and has no chance of being promoted anyway. Likewise, they do not compensate a person who shows hardly any effort in his or her life but grew up in poverty with few opportunities and incentives to work hard. In both cases, the choice not to work hard legitimizes a highly unequal outcome. While respondents view the unequal circumstances as unfair (discrimination: 81%, poverty: 73%), many consider the unequal outcomes of the vignettes as fair (discrimination: 96%, poverty: 82%).

Taken together, my findings suggest that the prevailing notion of meritocratic fairness is "*shallow*". Meritocratic fairness holds that individuals should not be judged by their external circumstances. However, people may not factor in the fact that these external circumstances also influence the choices that agents make. They still hold them responsible for choices made in unequal environments.[1] As a result, disadvantaged agents do not face a benefit but rather a "burden of the doubt". In the real-world, their counterfactual choices are almost always uncertain. And since they cannot verify what they would have done under better circumstances, they are judged by their actual choices, even when these are disadvantageously shaped by unequal circumstances.

These fairness views matter. They are likely to affect which inequalities people accept at the workplace (Akerlof and Yellen, 1990; Breza et al., 2018), they could shape hiring decisions, promotions, or college admissions, and affect which socioeconomic policies people support. For example, shallow meritocracy can *doubly* disadvantage the disadvantaged. Not only do they face adverse and discouraging circumstances, but they are also blamed and held responsible if they show less effort, dedication, and perseverance

[1] As a side note, valuable talents, traits, and abilities such as cognitive skills are also commonly viewed as important components of merit. However, these skills are also shaped by external circumstances (e.g., Alan and Ertac, 2018; Heckman, 2006; Kosse et al., 2019; Putnam, 2016), so a similar question arises for the effect of circumstances on skills. This study focuses on the effect of circumstances on choices because it is the simpler, more transparent, and relatable channel.

under these conditions. Moreover, affirmative action and redistributive policies, which aim to correct for this double disadvantage, are highly contentious and often opposed precisely because they are considered to be violating meritocratic fairness.

Related literature The study builds on and contributes to several strands of the literature. The fairness views of the general population have long been a focus of economic research because they are recognized as an important determinant of welfare systems and a defining feature of political culture (Alesina and Glaeser, 2004; Alesina and Angeletos, 2005; Alesina et al., 2018; Andreoni et al., 2020; Bonomi et al., 2021; Fisman et al., 2020; Gethin et al., 2021; Giuliano and Spilimbergo, 2013; Hvidberg et al., 2022; Kuziemko et al., 2015; Stantcheva, 2021). Past research documents that the idea of merit is at the center of fairness and inequality acceptance. Merit is associated with choices such as working hard or taking risks. Unequal rewards derived from unequally meritorious choices are typically considered fair and legitimate (Akbaş et al., 2019; Almås et al., 2020; Cappelen et al., 2007, 2010, 2013; Konow, 2000; Krawczyk, 2010; Mollerstrom et al., 2015), and even small differences in merit can sometimes justify large reward inequalities (Bartling et al., 2018; Cappelen et al., 2020a). Cappelen et al. (2022) show that even degenerate choices can have a meritorious character. Participants in their study reward "choices" even when the agents have no real choice and can only decide between two identical alternatives. Thus, choices appear to be central to merit judgments. But choices are always the result of both internal causes—an agent's type, their personality, or taste for hard work—and external causes, namely the ubiquitous effect of circumstances on choices. This study is the first to show that merit judgments do not noticeably differentiate between internal and external causes of choice, and it provides an in-depth analysis of the underlying behavioral mechanisms.

The finding that people are held responsible for their choices even if these choices are the product of external circumstances also relates to the literature on moral responsibility and moral luck (Baron and Hershey, 1988; Bartling and Fischbacher, 2012; Brownback and Kuhn, 2019; Cappelen et al., 2022; Falk et al., 2021, 2020b; Gurdal et al., 2013; Nagel, 1979). Individuals are often judged not only by their choices but also the consequences of their choices, even if these are accidental, unintended, and the product of chance. Here, I show that individuals can be held responsible for external luck not only if it shapes the consequences of their decisions but also when it directly impacts their decisions.

This study also connects to a recent literature on inference in economics (e.g., Benjamin, 2019; Bushong and Gagnon-Bartsch, 2022; Enke and Zimmermann, 2017; Graeber, 2022; Han et al., 2022; Liang, 2021). In particular, individuals often struggle with complex decisions in uncertain and contingent environments (Esponda and Vespa, 2014, 2019; Martínez-Marquina et al., 2019; Oprea, 2022)—a key element of counterfactual thinking. However, counterfactual thinking itself remains relatively unexplored in economics, although cognitive scientists have long recognized its centrality to causal reasoning and inference (Byrne, 2016; Engl, 2018; Kahneman and Miller, 1986; Lagnado and Gerstenberg, 2017; Roese, 1997; Sloman, 2005). This study illustrates that the inherent uncertainty of the counterfactual strongly affects individuals' judgments even though they accurately anticipate the expected counterfactual.

Finally, understanding the practice of meritocratic fairness informs the debate about the merits and myths of meritocracy led by social scientists and philosophers (Frank, 2016; Greenfield, 2011; Markovits, 2019; Sandel, 2020; Wooldridge, 2021; Young, 1958). This paper's contribution to the debate is to document the prevailing notion of meritocratic fairness: Choices are a critical determinant of merit, even when external circumstances influence which choices are made.

The remainder of this paper is structured as follows. Section 2 sets the stage with a brief conceptual discussion, Section 3 describes the main experimental design, and Section 4 presents the main results. Section 5 examines their behavioral foundations, and Section 6 reports the vignette study. Finally, Section 7 concludes the paper.

2 Conceptual discussion

The goal of the paper is to explore whether people's merit judgments take into consideration that others' choices are often substantially shaped by circumstances. To fix ideas, this section discusses and compares two conflicting meritocratic fairness views that people could endorse.

As a motivating example, consider the following case of racial discrimination in the labor market. A white employee and an employee of color can choose whether to work hard for a promotion. However, their boss is notorious for being racist and has never promoted employees of color before. The white employee decides to work hard to win the promotion, the employee of color does not. In the end, the white employee is promoted and awarded an attractive bonus, while the employee of color is not.

When judging whether the outcome of this illustrative story is fair, two intuitions collide. On the one hand, the white employee has worked harder, so he or she might deserve the promotion and the bonus. On the other hand, their effort choices have been shaped by the highly unequal and unfair circumstances of racial discrimination. This simple story captures the essence of a fundamental question for meritocracy. If we want to reward others according to their effort choices but not their circumstances, do we hold them responsible for their choices when these choices are shaped by unequal circumstances?

More generally, consider a situation where two workers choose how much effort to exert, but unequal circumstances encourage one of the workers to work hard, while they discourage the other worker. I distinguish between two meritocratic views on how merit in such a setting should be evaluated, which I refer to as "actual choice meritocratism" and "comparable choice meritocratism".

Actual choice meritocrats hold people fully responsible for their choices, even if these choices are shaped by unequal external circumstances. Their merit judgments comove with the effort choices that workers make. Whether these choices result from different environments is considered irrelevant. This view often seems to underlie the public debate where the idea that people should be held responsible for their bad choices—be it in school (laziness, misdemeanor), health (nutrition, smoking), or at work (low career ambitions, low effort)—is paramount, often without regard to individuals' circumstances (see Greenfield, 2011, for a discussion).

By contrast, *comparable choice meritocrats* do not hold individuals responsible for external causes of choice but only for internal causes.[2] In economics, this view has been prominently endorsed by Roemer (1993). Roemer argues that if individuals cannot be held responsible for their circumstances, they are also not responsible if these circumstances induce poor choices. Hence, when circumstances influence effort, merit and raw effort cannot be equated. Instead, merit judgments need to correct for external influence on choice. Comparable choice meritocrats, therefore, want to compensate workers for any discouraging situational influence. One option to account for this is to ask which choices the workers would have made in a fully comparable situation. For example, they could ask how hard the disadvantaged worker would work if his returns to effort would also be high. Then, they base their merit judgment on this counterfactual, comparable effort choice.[3] Of course, this requires an inference about counterfactual comparable choices, which, if biased, could prevent comparable choice meritocrats from consistently applying their fairness view.

Conceptually, there are intriguing *normative* arguments for both actual choice and comparable choice meritocratism.[4] Here, however, the research question is of *positive*

[2]These internal causes of choice, such as type or preference differences, can often be attributed to differential external circumstances as well—be it nature or nurture (Cesarini et al., 2009; Dohmen et al., 2012; Harden, 2021; Heckman, 2006; Kosse et al., 2019). While outside the scope of this paper, one could hence even ask whether these differences can justify merit differences.

[3]In principle, comparable choice meritocrats could also base their merit judgments on counterfactual effort choices in another environment, e.g., low returns to effort. Relatedly, Roemer (1993) takes an individual's relative ranking in the effort distribution conditional on circumstances as a comparable measure of merit. These details affect neither the qualitative argument here nor the interpretation of later treatment effects.

[4]For instance, incentives to behave well could deteriorate if individuals are not fully accountable for their actual choices. Moreover, workers already bore the costs of their working decisions. Why should a lazy worker be rewarded for the hard work he would have done (but did not do) in a counterfactual environment? On the other hand, it seems inconsistent to claim that external circumstances should not influence merit judgments, while their external influence on choice does.

nature. The study investigates which merit judgments the general population makes. First, are they sensitive to the effect of circumstances on choices? Second, if not, are they insensitive because comparable choice meritocrats are absent from the population or because they incorrectly infer what would have happened under equal circumstances and fail to apply their merit view?

3 Experimental design

Studying how the effect of circumstances on effort choices shapes merit judgments requires a setting where choices are central to merit and merit judgments can be measured in an incentivized way. And it requires experimental conditions that exogenously vary how circumstances affect choices. Below, I describe how I tailor the experimental design to meet both requirements.

3.1 Setting: Merit judgments

I create an experimentally controlled situation of inequality between *workers* (referred to as "he") and observe how study participants (*spectators*, referred to as "she") redistribute money between workers, conditional on workers' effort choices. Spectators decide which reward each worker deserves and thereby judge which merit originates from the workers' choices.

Workers I hire US workers on Amazon's online labor market Mechanical Turk for a crowd-working job in which they collect email address data for another research project. In each task, a worker is given the name of a person, searches for the person's website, identifies their email address, and enters it in a data collection form. Typically, it takes about two minutes to complete one task. The crowd-working job does not require special qualification but demands effort and time, ensuring that hard work rather than skill determines success. Each worker k earns a piece rate π_k (his returns to effort) and can freely choose how many tasks E_k to complete. Workers know that a lottery determines their piece rate, which can either be high ($0.50) or low ($0.10). The initial payment of a worker is $\pi_k E_k$. Workers know that someone else might influence their payment, but they neither know when, why, nor how this happens, nor who is involved in this process. This guarantees that workers cannot distort their effort decisions in anticipation of a later redistribution stage. Each worker additionally receives a fixed remuneration of $1. The instructions for workers are available online (https://osf.io/xj7vc/).

For the redistribution stage, workers are assigned to pairs. I will refer to the two workers in a pair as worker A and worker B. I focus on pairs where worker A receives a high piece rate of \$0.50 and worker B receives a low piece rate of \$0.10.[5] Inequality between the two workers is likely to prevail—either due to differences in effort E_k or the piece rate π_k. Whereas effort E_k is a choice variable, the piece rate π_k is outside the control of workers but is likely to shape the workers' effort choices. Indeed, workers complete, on average, more than three times as many tasks (mean: 16.8 tasks) for a high piece rate of \$0.50 than for a low piece rate of \$0.10 (mean: 5.0 tasks, see Appendix E), rendering the setting well-suited to study whether merit judgments take into account that circumstances affect workers' choices.

Spectators I invite adults from the general US population to participate in the online experiment. Each study participant ("spectator") is assigned to a pair of workers and is informed about the task, situation, choices, and earnings of the workers. In particular, spectators know that a lottery determines the workers' piece rate. Spectators then determine the final earnings of both workers and judge which percentage share of the total performance-based earnings each worker deserves. That is, they can redistribute the earnings between both workers. Redistribution comes at no cost.[6] Spectators know that their decision is strictly anonymous and that workers are unaware of the redistribution stage. I implement the merit judgments of 100 randomly selected spectators so that spectator decisions are (probabilistically) incentivized.[7] Their decisions can have real and meaningful consequences for workers. Appendix G provides the main instructions for spectators, and the complete instructions are available online (https://osf.io/xj7vc/).

The redistribution decisions of spectators, neutral third parties who have no monetary stake in the distribution of funds, commonly serve as a measure of fairness behavior and views (e.g., Almås et al., 2020; Andreoni et al., 2020; Cappelen et al., 2013; Konow, 2000; Mollerstrom et al., 2015). They mirror the fact that society's fairness views are often implemented via redistributive schemes that intervene into naturally arising market outcomes. To elicit spectators' merit judgments for various effort choices, I employ a contingent response method. Each spectator decides whether and how to redistribute the earnings in eight different effort scenarios. Each scenario describes how many tasks worker A and how many tasks worker B completed. The first seven scenarios are hypo-

[5] In the experiment, I randomly vary whether worker A or worker B is the worker with the advantageous, high piece rate. Reassuringly, I find that this variation is irrelevant for spectators' redistributive behavior. Here, I recode all responses as if worker A were the advantaged worker to ease analysis and exposition. Furthermore, sometimes both workers in a pair receive a piece rate of \$0.10 or both receive a piece rate of \$0.50. These worker pairs are used in additional experimental conditions that I will introduce later.

[6] I abstract from the frequently studied fairness-efficiency trade-off. Existing research shows that fairness concerns often dominate efficiency concerns (Almås et al., 2020). Spectators cannot redistribute the fixed remuneration of \$1 but only the performance-based rewards.

[7] Charness et al. (2016) review the advantages and disadvantages of implementing the decisions of a subset of participants versus those of all participants. The literature documents little difference between both methods for the estimation of treatment effects.

thetical, presented in random order, and selected to represent various effort shares of worker B (denoted by $e = \frac{E_B}{E_A+E_B}$). Panel A of Table 1 summarizes these effort scenarios. For example, in scenario 1, worker A does all the work and completes 50 tasks, while worker B does not complete any task ($e = 0\%$). In scenario 4, both workers complete 25 tasks ($e = 50\%$). Furthermore, in scenario 7, worker A completes 0 tasks and worker B completes 50 tasks ($e = 100\%$). The other scenarios present intermediate cases. The eighth scenario is real and describes how many tasks the two workers actually complete. Spectators' decisions in this scenario determine the final payoff of the workers. However, spectators are not told which scenario is real and therefore have to take each of their decisions seriously.[8]

Effort choices in the real scenario vary across experimental conditions (introduced in the next subsection) due to the incentive effects of the conditions. Thus, the real scenario does not allow for a consistent comparison across treatments. To avoid this problem, I only analyze the merit judgments in the first seven scenarios. The contingent response method is important for the identification because it allows analyzing merit judgments for the same effort scenario and effort choices across the treatment and control conditions.

3.2 Conditions: Varying the effect of circumstances on choices

In a between-subject design, I exogenously vary whether the effort choices of workers are differentially affected by circumstances. For this purpose, I manipulate *when* the workers learn about the realized piece rate of their lottery and inform the spectators about this. Panel B of Table 1 provides an overview of both conditions.

> **Control condition**: Both workers do not know their realized piece rate while making their effort choices. They are aware that their piece rates might either be \$0.50 or \$0.10 with equal chance. They learn about their realized piece rate (\$0.50 for worker A and \$0.10 for worker B) only after they finish their work.
>
> **Treatment condition**: Both workers are informed about their realized piece rate already before they decide how much effort they exert. Thus, worker A knows about his high rate of \$0.50 and worker B about his low rate of \$0.10 when they decide how many tasks they complete.

The experimental conditions vary whether the two workers in a pair optimize against identical or different piece-rate expectations. In the control condition, both workers

[8] The contingent response method requires that spectators believe that each scenario is potentially true. I confirm this in a series of additional analyses described in Appendix B.2. In short: (i) few spectators can distinguish the hypothetical scenarios from the real one; (ii) the results are robust to excluding respondents who recognize the real scenario; and (iii) the results can be replicated in a robustness experiment that does not employ the contingent response method.

Table 1 Overview of effort scenarios and experimental conditions

(A) Effort scenarios (presented in random order)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Effort share of worker B: e	**0%**	**10%**	**30%**	**50%**	**70%**	**90%**	**100%**
Effort of worker A	50	45	35	25	15	5	0
Effort of worker B	0	5	15	25	35	45	50
Payment of worker A	\$25.00	\$22.50	\$17.50	\$12.50	\$7.50	\$2.50	\$0.00
(Share)	(100%)	(98%)	(92%)	(83%)	(68%)	(36%)	(0%)
Payment of worker B	\$0.00	\$0.50	\$1.50	\$2.50	\$3.50	\$4.50	\$5.00
(Share)	(0%)	(2%)	(8%)	(17%)	(32%)	(64%)	(100%)

(B) Experimental conditions (between-subject)

	Control condition		*Treatment condition*	
Worker	**A**	**B**	**A**	**B**
Constant across conditions				
Realized π	\$0.50	\$0.10	\$0.50	\$0.10
Effort choices	*Depends on effort scenario*			
Payment	*Results from effort scenario and realized π*			
Varies across conditions				
Expected π	\$0.50 or \$0.10 each with 50%	\$0.50 or \$0.10 each with 50%	\$0.50	\$0.10

Notes: Panel A presents an overview of all effort scenarios. Panel B summarizes and compares the experimental conditions.

face the same expected circumstances and respond to the same environment so that their effort choices are comparable. If one worker completes more tasks, this directly signals his higher baseline willingness to work hard. In the treatment condition, the workers face different circumstances and their effort choices are differentially shaped by circumstances. The high piece rate encourages worker A to work more, whereas the low piece rate discourages worker B. Thus, if the advantaged worker A completes more tasks, this also reflects advantageous circumstances. By comparing spectators' redistributive behavior across treatment and control, I can test whether and how merit judgments react to the effect of circumstances on choices.

The contingent response method allows me to study merit judgments and their sensitivity to circumstances' effect on choices in seven different effort scenarios. Each scenario describes how much effort each worker exerts and how much money they initially earn. The scenarios are identical across the treatment and control conditions, but their interpretation changes. For instance, two workers who complete 25 tasks each (scenario 4) show identical diligence in the control condition. However, in the treatment condition, working on 25 tasks for a \$0.50 piece rate signals a much lower baseline willingness to work hard than working on 25 tasks for a \$0.10 piece rate. As another example, if worker A completes 50 tasks and worker B does nothing (scenario 7), worker

A clearly signals higher diligence in the control condition. The situation is less clear in the treatment condition because the effort choices can be partially attributed to unequal circumstances.

For actual choice meritocrats, the difference between the treatment and control condition is irrelevant. Their merit judgments depend solely on workers' actual effort choices, which are identical across both conditions. But comparable choice meritocrats who recognize that worker B is discouraged by his circumstances and would likely work harder for a high piece rate should compensate him with a higher reward share.

3.3 Experimental procedures

Workers I recruited 336 workers on Amazon Mechanical Turk in May and June 2020 to participate in the crowd-working job. On average, workers complete 12 tasks and earn about $5.40. I form 100 pairs with 200 of those workers and use them to incentivize the redistribution decisions of spectators.[9]

Spectators I recruit a sample of 653 participants in collaboration with Lucid, an online panel provider that is frequently used in social science research (Coppock and McClellan, 2019; Haaland et al., 2022). The sample excludes participants who do not complete the first seven redistribution decisions or speed through the experimental instructions (see Appendix A). The sampling plan and the exclusion criteria were preregistered (see Appendix F). Participants are broadly representative of the adult population of the US in terms of gender, age, region, income, and education. Table 2 displays summary statistics from the sample and compares them to the data obtained from the American Community Survey 2019. The sample closely follows the characteristics of the American population, except perhaps for education: 43% of the sample possess an undergraduate degree, compared to about 31% of the US population. Respondents were randomly assigned to either the treatment ($n = 329$) or the control ($n = 324$) condition. Appendix Table A.3 shows that the covariates are balanced across experimental conditions.

The experiment was conducted online in June 2020. Most participants spent 10 to 30 minutes to complete the experiment (15% and 85% percentile), with a median response duration of 16 minutes. The experiment is structured as follows. First, participants answer a series of demographic questions that monitor the sampling process. Inattentive participants are screened out in an attention check. Detailed instructions on the workers' situation and the redistribution decisions follow. The experimental treatment-control variation is introduced only at the end of the instructions. This guar-

[9] I ran the main experimental conditions together with additional robustness and mechanism conditions with a total of 1,855 participants. The additional conditions will be introduced later. The workers were recruited jointly for all experimental conditions. Appendix A provides an overview. Workers who were not selected for the redistribution stage received their original performance-based payments.

Table 2 Comparison of the sample to the American Community Survey

Variable	ACS (2019)	Sample
Gender		
Female	51%	51%
Age		
18-34	30%	30%
35-54	32%	33%
55+	38%	37%
Household net income		
Below 50k	37%	40%
50k-100k	31%	34%
Above 100k	31%	27%
Education		
Bachelor's degree or more	31%	43%
Region		
Northeast	17%	21%
Midwest	21%	21%
South	38%	36%
West	24%	22%
Sample size	2,059,945	653

Notes: Column 1 presents data from the American Community Survey (ACS) 2019. Column 2 presents data from the representative online sample.

antees that the instructions about the workers' task and the redistribution decisions are understood and interpreted identically across conditions. Then, a quiz tests whether participants understand the key aspects of the experiment and corrects them if necessary. Subsequently, participants make their redistribution decisions. Each redistribution decision screen also contains a tabular summary of the workers' situation, including their expected and realized piece rates, to ensure that this information is salient in the moment of decision making. Finally, a series of follow-up questions are asked to collect additional demographic variables and investigate possible mechanisms. Respondents also explain in an open-text format what thoughts and considerations shaped the merit judgments they made.

3.4 Additional experiments

I run a series of additional conditions and experiments to explore the robustness of the results and shed light on their behavioral mechanisms. The details will be introduced in later sections. Table 3 provides an overview and brief description of all conditions and studies.

Table 3 Overview of all experiments

Study	Description
Main study	Varies whether unequal circumstances encourage/discourage effort.
Robustness	
“Equal rates” conditions*	Replicate main study, but workers receive same piece rate.
Disappointment study	Explores motive to compensate workers for disappointment.
Leisure time study	Replicates main study, but workers choose work or leisure time.
Mechanism	
Attention condition*	Shifts attention towards the effect of circumstances on choices.
“Equal rates” attention cond.*	“Equal rates” version of the attention condition.
Counterfactual study	Reveals what would have happened in equal circumstances.
Advantaged counterfactual study	Replicates counterfactual study, but reveals counterfactual choice of advantaged (instead of disadvantaged) worker.
Rationale study	Asks spectators to explain their reward decision.
External validity	
Vignette study	Explores merit judgments in real-world scenarios.
Vignette evaluation study	Participants evaluate fairness of unequal outcomes.
*Run in parallel to main study.	

Notes: This table lists all studies that I present in this paper. Only the main study is introduced in this section. The details of all other experimental conditions and studies will be introduced in later sections.

4 Main result

4.1 Merit judgments in the control condition

I start by studying the merit judgments of spectators in the control treatment to first understand how they distribute rewards when effort choices are still comparable. In the control condition, workers make their effort choices in an identical environment: both workers expect either a $0.50 or $0.10 piece rate (each with 50%). Only after completing their work, worker A learns that he randomly receives the high piece rate of $0.50, whereas worker B learns that he earns $0.10 per completed task. Do spectators compensate worker B for the bad luck of a low piece rate?

Figure 1 visualizes the average share of the total earnings that spectators assign to the disadvantaged worker B. Panel A displays the mean share, averaged across all seven scenarios, and Panel B presents the results in each of the seven effort scenarios. The results show that spectators indeed counterbalance the bad luck of a low piece rate. They strongly redistribute money from worker A (high piece rate) to worker B (low piece rate). Averaged across scenarios, worker B receives 44.1% of the total earnings (red bar), which is much higher than the share he would receive without redistribution (31.9%, gray line). A similar picture emerges across the different effort scenarios depicted in Panel B.

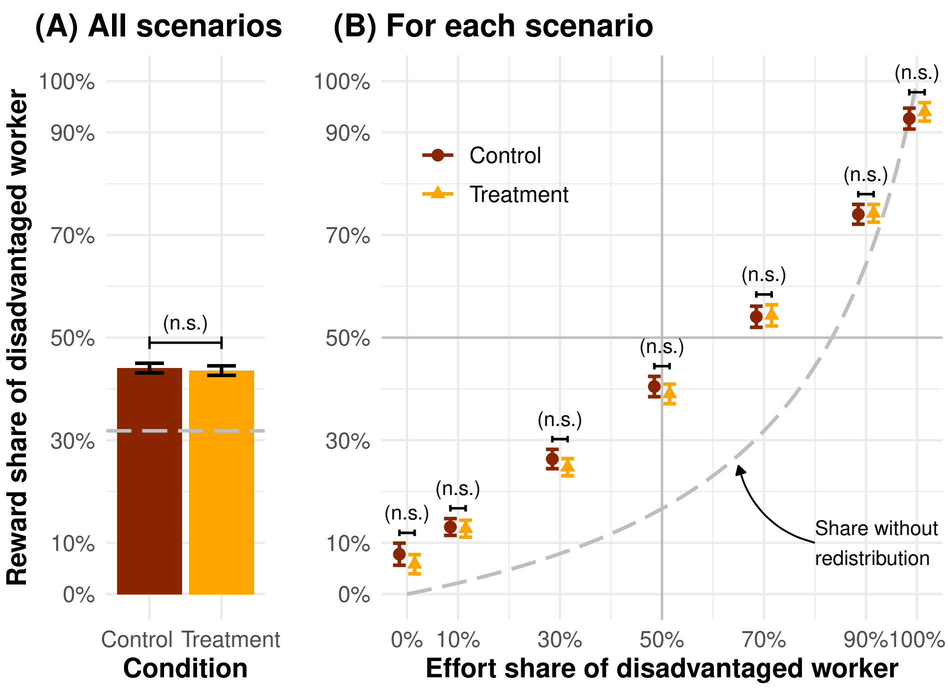


Figure 1 Main experiment: Mean reward share of disadvantaged worker (95% CI)

Notes: Results from the main study. Panel A displays the mean reward share assigned to the disadvantaged worker B in both experimental conditions, averaged across all seven effort scenarios, with 95% confidence intervals. Panel B plots the mean reward share in each effort scenario with 95% confidence intervals. The gray dashed line shows the default share, that is, which payment share worker B would receive if spectators do not redistribute. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, (n.s.) $p \geq 0.10$.

However, this compensation is not unconditional; rather, it strongly depends on the effort choices that workers make. In fact, many participants reward worker B proportionally to his effort share, whereby, they assign him a payment share that is equal to the share of the tasks he completes (Appendix Figure B.1). As a result, the average reward share assigned to the disadvantaged worker moves closely with his effort share. For example, the disadvantaged worker receives only 8% if he completes 0% of the tasks, but 26% if he completes 30% of the tasks, 40% if he completes 50% of the tasks, or 74% if he completes 90% of the tasks. Deviations from a reward distribution based purely on effort indicate traces of libertarian and egalitarian redistributive behavior. A small share of "libertarian" spectators never redistribute and always accept the pre-existing reward shares, and a small share of "egalitarian" spectators always implement equal shares irrespective of the workers' effort decisions (see Figure B.1). I will revisit this heterogeneity in fairness views in Section 5.4. For now, the key conclusion is that, in the control condition where both workers react to the same environment, merit mostly derives from effort choices.

4.2 Treatment effect on merit judgments

This sets the stage for my main research question. Spectators reward workers for effort but compensate them for unequal circumstances. But do they also take into account that circumstances can shape workers' effort choices? To find out, the treatment condition informs workers about their realized piece rates already before they make their effort choice. Consequently, worker B is additionally disadvantaged: he is discouraged to work hard by a low piece rate of $0.10. By contrast, worker A is encouraged by a high piece rate of $0.50. I test whether spectators assign a higher reward share to worker B in the treatment than in the control condition to compensate him for this disadvantage.

The results show that merit judgments are insensitive to the effect of circumstances on choices. Figure 1 shows that the payment shares are virtually indistinguishable between the treatment and the control condition. Worker B receives on average 43.6% of the total earnings in the treatment condition and 44.1% in the control condition (Panel A). Therefore, spectators do not compensate worker B for the disadvantageous and discouraging effect of a low piece rate on effort choices in the treatment condition. They even assign him an (insignificant) 0.49 percentage points (pp) lower share ($p =$ 0.464; see Table 4). Panel B shows that this conclusion holds for all seven scenarios. Irrespective of whether worker A or B completes more tasks, or both work equally hard, spectators do not counterbalance the effect of circumstances on choices. None of the seven treatment-control comparisons detects a significant difference, nor does a highly powered joint F-test that tests the null hypothesis that treatment differences are zero in all seven effort scenarios ($p = 0.668$).[10]

This null result does not reflect a noisy estimate but rather constitutes a precisely estimated null finding.[11] Due to a sample of 653 individuals and 4,571 decisions, the study is well powered and possesses a minimum detectable effect size for the treatment effect averaged across scenarios of about 2 pp (at 80% power). The 95% confidence interval of the treatment effect ranges from -1.8 to 0.8 pp. This means that I can reject even tiny effect sizes with high statistical confidence, namely that workers who are disadvantaged by circumstances' effect on choices receive a compensation of more than 0.8 pp of the total payment. Two successful replications reported below corroborate this assessment.[12]

[10]The F-test is derived from a regression of worker B's payment share r_{is} on a treatment dummy interacted with a dummy for each scenario s and scenario fixed effects. It tests the null hypotheses that the treatment effects are zero in all seven effort scenarios. Standard errors are clustered at the participant level.

[11]Precisely estimated null results are very informative from a Bayesian learning perspective—often even more informative than rejections of a null hypothesis (Abadie, 2020).

[12]Even if I focus on the treatment effect among meritocratic spectators who distribute rewards conditional on effort, the study remains highly powered. In Section 5.4, I estimate that approximately 70% of the US population assign rewards meritocratically. Hence, assuming that all other spectators do not react to the treatment difference, the minimum detectable effect size and the width of confidence intervals for

Table 4 Treatment effects on the average reward share of disadvantaged worker

	Mean reward share of disadvantaged worker (in %)				
	Main	Robust: No quiz mistakes	Robust: Decisions 1-3	Robust: High duration	Robust: With controls
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.493	-1.002	-0.135	0.160	-0.353
	(0.673)	(0.827)	(1.335)	(0.785)	(0.684)
Constant	44.068***	44.792***	43.652***	43.479***	47.264***
	(0.480)	(0.573)	(0.915)	(0.553)	(4.569)
Controls	–	–	–	–	✓
Observations	653	395	653	471	634
R^2	0.001	0.004	0.000	0.000	0.004

Notes: Results from the main study, ordinary least squares (OLS) regressions, robust standard errors in parentheses. The outcome variable is the reward share (in %) a spectator assigns to the disadvantaged worker B, averaged across all seven effort scenarios. The independent variable is a treatment indicator. Column 1 presents the main specification. Columns 2-5 present different robustness specifications: Column (2) excludes respondents who initially answer at least one quiz question incorrectly, Column (3) considers only the first three decisions of each participant, Column (4) excludes the 25% respondents with the lowest response duration, and Column (5) includes controls (indicators for female gender, college degree, and being Republican, as well as log income, and age). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

An average null effect, even if precisely estimated, could still conceal meaningful treatment effects for parts of the population. Therefore, I test for heterogeneous treatment effects. In a first step, I test for heterogeneity alongside six preregistered covariates: gender, education, party affiliation, income, empathy, and internal locus of control. I assess empathy with four survey questions that measure perspective taking and empathetic concern adopted from Davis (1983) and locus of control with a streamlined four-item scale developed in Kovaleva (2012). An internal locus of control measures whether a person attributes successes and failures to his or her own actions and abilities rather than attributing them to luck, fate, or the actions of others. None of these variables significantly moderates the treatment effect, and none is significantly associated with merit judgments in the baseline control condition (see Table B.2). In a second step, I apply the model-free approach of Ding et al. (2016) that tests whether *any* significant treatment heterogeneity exists. The method relies on randomization inference and basically tests whether the treatment distribution of the outcome variable is identical to the control distribution shifted by the average treatment effect. No significant heterogeneity in treatment effects is detected ($p = 0.446$).

Together, the results provide strong evidence for the absence of a meaningful effect.

Result 1: Individual merit judgments do not factor in that circumstances influence

an effect *among meritocratic spectators* increases by a factor of $1/0.7 \approx 1.4$ and thus remains reasonably small.

choices. People reward others based on their effort, even if effort choices are unequally shaped by external circumstances.

I interpret the evidence as a "proof of concept". Merit judgments can be blind to the circumstances shaping the choices that others make, even in a simple, transparent setting in which effort choices and circumstances are perfectly observed. This insensitivity to the effect of circumstances on choices likely characterizes merit judgments in many other contexts—an issue that I explore empirically in Section 6. First, however, the advantages of the controlled experimental set-up are utilized to address possible robustness concerns and explore the behavioral mechanisms.

4.3 Robustness

I replicate the main result in multiple robustness checks.

Noisy responses In the first set of robustness tests, I ensure that the findings are not driven by a misunderstanding of the instructions, survey-taking fatigue, or inattentive participants—all of which would increase survey noise and thus could potentially conceal treatment effects. In Column 2 of Table 4, I exclude participants who initially answer one of the control questions incorrectly, which could indicate a lack of understanding. In Column 3, I restrict the analysis to the first three redistribution decisions each participant makes, which would arguably be less affected by survey fatigue. In Column 4, I exclude the 25% of participants with the lowest response duration to drop participants who might "speed through" the survey. All three specifications replicate the main results. Moreover, I obtain virtually identical results if I control for respondents' demographic background (Column 5).

Open-text responses Second, I corroborate my main findings by analyzing the open-text responses in which participants explain how and why they made their merit judgments. For the analysis, each response is manually classified into different fairness arguments (see Appendix B.3). Almost no participant in the treatment condition (1%) mentions that they consider that workers' choices are shaped by circumstances. By contrast, most participants (59%) argue that workers' effort choice should determine the final payments. Furthermore, the explanations offered by the respondents do not significantly differ between the treatment and control condition (see Table B.5), thus replicating the main findings.

Salience of direct effect of unequal piece rates Third, one might be concerned that the direct effect of the piece rates on earnings is too salient and crowds out attention to the effect of circumstances on choices. For example, a disadvantaged worker who completes 15 tasks and earns only $1.50 would have earned $7.50 with a high piece rate. Spectators might primarily think about this difference and therefore overlook that the worker would also have worked much harder (e.g., complete 35 tasks for a payment of \$17.50). However, evidence from two additional experimental conditions that I ran in parallel to the main study does not support this explanation ("equal rates" conditions, $n = 661$, Appendix B.4). The two conditions keep the realized piece rate of both workers constant. In the control "equal rates" condition—analogously to the main experiment—both workers have identical expectations about their piece rate (\$0.10 or \$0.50 with an equal chance). In the treatment "equal rates" condition, worker A expects to earn either \$0.50 or \$0.90, whereas worker B expects to earn only \$0.10 or \$0.50. Thus, worker A is advantaged and encouraged to work hard, whereas worker B is disadvantaged and discouraged from working hard. However, in both conditions, chance determines that both workers earn the same rate of \$0.50, so that their initial earnings are fully proportional to their effort. Consequently, there is no direct piece-rate effect on payments that could distract spectators. Nonetheless, this independent robustness experiment fully replicates the main results. I do not detect significant differences in merit judgments across the two conditions. Again, the null result is obtained with high precision (Appendix Table B.6).

Compensation for disappointment Fourth, another potential concern is that a compensation for disappointment confounds the null effect. Worker B receives bad news upon learning that he only earns a low piece rate, and the timing of bad news could matter. In the control condition, worker B receives this information only after he stopped working, which could lead to greater disappointment. If spectators share this concern, they might want to assign a higher payment share to worker B in the *control* condition to compensate him for the higher disappointment. Any such effect would run opposite to the main treatment effect and could therefore conceal its existence if, by chance, the two effects offset each other in all seven effort scenarios. To be on the safe side, I design an additional experiment that rules out this confounding channel (disappointment study, $n = 606$, run in February 2021 with a US convenience sample, Appendix B.5). I replicate the main design with one crucial exception: Workers do not have a choice. Instead, all workers have to complete exactly ten tasks. Since no choice is involved, unequal circumstances cannot shape effort choices, and there is no reason to compensate for it. However, the motive to compensate for the timing of bad news is still present. If it matters, spectators should compensate worker B with a higher payment share in the control condition. The results reveal a negligible and insignificant difference that could not even conceal a minor treatment effect (Appendix Table B.7).

Work versus leisure time Fifth, a drawback of the naturalistic working context, which allows workers to quit the experiment once they stop working on the tasks, is that I cannot fix spectators' beliefs about what workers do afterwards. Some spectators might think that workers use the freed-up time to work and earn money on other online jobs. Although this consideration applies equally to the control and treatment conditions, it could obscure the meaning of fair rewards within the context of my study. Therefore, I conducted a final robustness experiment, the "leisure time" study, with 1,095 spectators whom I recruited from the US population via the survey platform Prolific in June 2022 (Appendix B.6). In the experiment, workers face the decision whether to enjoy leisure time for 30 minutes (watch videos on YouTube) or work for 30 minutes (collect email addresses).[13] In either case, they spend 30 minutes. As in the main study, a lottery determines whether they earn a high reward (here: £5 with 50% chance) or a low reward (here: £1 with 50% chance) for 30 minutes of work.[14] In the control condition, workers only learn about their realized rewards after they completed their work/leisure time, while the treatment condition informs them before they make their work/leisure choice. Once again, I find that the effort choices of workers strongly respond to their circumstances. Workers who know that they can earn £5 are approximately three times more likely to work than those who know that they can earn only £1. However, I find no quantitatively important treatment effect on the merit judgments of spectators. Pooled across effort scenarios, the treatment only increases the reward share of the disadvantaged worker by a non-significant 2.0 pp with an estimated standard error of 1.4 pp (Appendix Table B.8).

A pooled test Finally, drawing on the data from the main study, the "equal rates" study, and the "leisure time" study, I can estimate a pooled test that combines data from 3 independent studies, 2,409 individuals, and more than 10,000 decisions. The pooled test has a minimum detectable effect size of 1 pp. I replicate the null result and estimate a treatment effect of 0.0 pp with a 95% confidence interval of -0.7 to 0.7 pp (Appendix Table B.9).

5 Mechanism

This section investigates why merit judgments are insensitive to the effect of circumstances on effort choices. The conceptual framework of Section 2 suggests two explanations. On the one hand, the effect of circumstances on choices could simply be irrelevant for merit views. Spectators' fairness preferences might hold that merit should be solely grounded on actual effort choices ("actual choice meritocratism"). On the other hand, spectators might actually prefer to correct for the effect of circumstances on choices ("comparable choice meritocratism"), but they struggle to do so because they fail to

[13]The binary nature of the working decision strongly simplifies the set of possible effort choice scenarios, which allows me to run the experiment without the contingent response method. Every spectator makes exactly one merit judgment for a pair of workers, conditional on workers' *real* choices.

[14]Prolific pays participants in British pound.

infer what would have happened in identical comparable circumstances. Here, I investigate three behavioral obstacles that could impair spectators' inference—the fundamental attribution error, a lack of attention, and the uncertainty of the counterfactual—and I explore the heterogeneity of spectators' fairness preferences.

5.1 Fundamental attribution error

Spectators might overly attribute choices to the decision maker and underestimate the role of circumstances, that is, that workers' effort strongly reacts to the rate that they earn. Such an inferential error would be in line with the so-called fundamental attribution error, namely the notion that individuals underestimate situational influences on human decisions (Ross, 1977). If spectators underestimate the effect of circumstances on the choices of workers, they have little reason to correct for it. To shed light on this mechanism, the main study elicits participants' beliefs about how workers' effort choices react to the piece rate. Spectators learn that workers complete on average five tasks for a \$0.10 piece rate and estimate how many tasks workers complete on average for a \$0.50 piece rate. Their responses are incentivized: One in ten participants earns a \$5 Amazon gift card if their response is at most one task away from the true value.

Results The findings are not consistent with a fundamental attribution error. On average, participants believe that workers complete 3.46 times as many tasks for a rate of \$0.50 than for a rate of \$0.10. Therefore, the perceived incentive effect is even slightly larger (though not significantly) than the observed effect of 3.33 (p = 0.749, t-test, Figure C.1). Moreover, even if I estimate the treatment effect only among spectators who believe in a strong effect of circumstances on choices, I do not find any sign of positive treatment effects (Appendix Table C.1).

5.2 Attention

Spectators could be unaware that circumstances shape effort choices *while* making their merit judgments. Once explicitly asked about it, participants acknowledge that the effect exists, but it might still escape their attention while making their merit judgments. Attention (or lack thereof) is a powerful explanation of behavior in many other domains (e.g., Andre et al., 2022; Chetty et al., 2009; Gabaix, 2019; Taubinsky and Rees-Jones, 2018). To test for this mechanism, I ran an additional experimental condition in parallel to the main study that draws the attention of participants to the effect of circumstances on effort choices just before their merit judgments (n = 274). As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix A).

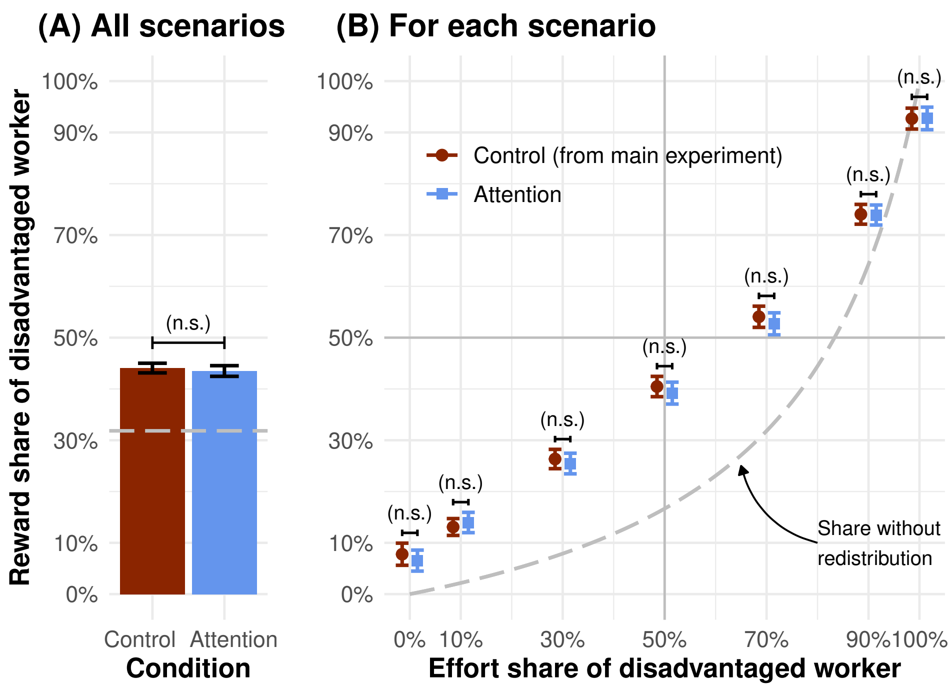


Figure 2 **Attention manipulation**: Mean reward share of disadv. worker (95% CI)

Notes: Results from the attention condition and the control condition of the main study. Panel A displays the mean reward share assigned to the disadvantaged worker B in both experimental conditions, averaged across all seven effort scenarios, with 95% confidence intervals. Panel B plots the mean reward share in each effort scenario with 95% confidence intervals. The gray dashed line shows the default share, that is, which payment share worker B would receive if spectators do not redistribute. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, (n.s.) $p \geq 0.10$.

> **Attention condition**: I explicitly inform spectators that "the piece rates strongly influence the number of tasks a worker completes." Spectators learn how large this incentive effect is on average and read two typical comments by workers that explain why this is the case. For example, the comment of a typical disadvantaged worker with a \$0.10 rate is: "For the amount of time that goes into these tasks, the compensation is simply just not sufficient." Participants have to spend at least 20 seconds on this information page, whose key message is repeated on the next page and tested for in the subsequent quiz.

Combining a qualitative statement, quantitative information, and workers' first-hand comments on their own experiences ensures that it is salient to spectators that circumstances shape choices. If a lack of attention explains the main result, spectators should compensate the disadvantaged worker with a higher reward share in the attention condition compared to the baseline control condition.

Results However, participants who are informed about and focused on the effect of circumstances on choices still do not compensate the disadvantaged workers. Figure 2 visualizes this result (following the format of Figure 1). As before, the null effect is precisely estimated and present in each of the seven effort scenarios (Panel B). Aggregated across scenarios, the mean payment share of worker B is 43.5% in the attention condition versus 44.1% in the control condition (Panel A). The 95% interval of their difference allows me to rule out even tiny treatment effects of 0.8 pp (Appendix Table C.2.A). I also find virtually no difference between the attention condition and the treatment condition of the main experiment (mean payment share: 43.6%, Appendix Table C.2.B).[15] Hence, a lack of attention to the effect of circumstances on effort choices cannot explain the results. Taking stock, I conclude:

> **Result 2**: Merit judgments do not factor in that circumstances influence choices, even though spectators are aware of and accurately anticipate the average effect of circumstances on choices.

5.3 Uncertainty of the counterfactual

Compensating worker B for the disadvantageous effect of circumstances on choices does not only require understanding and awareness of the average effect of circumstances. It also raises the question of what the two specific workers to whom a spectator has been assigned would have done under identical circumstances. How many tasks would worker B have completed had he also earned a high piece rate of $0.50? Such a counterfactual benchmark would underlie the reward decision of a comparable choice meritocrat who believes that choices in comparable circumstances should form the basis of merit judgments. However, this counterfactual is unknown and uncertain, even for spectators who accurately anticipate the average effect of circumstances on choices. Recent research shows that people struggle with complex decisions in uncertain and contingent environments, rendering this a promising explanation for why the merit judgments of spectators are insensitive to the effect of circumstances on choices (Esponda and Vespa, 2019; Martínez-Marquina et al., 2019; Oprea, 2022). Spectators might abstain from any conjecture and base their merit judgments on what they know with certainty: observed effort levels.[16]

I devise a new mechanism experiment in which some spectators are explicitly informed about worker B's counterfactual effort choice, thereby removing any uncertainty about the counterfactual state (counterfactual study, $n = 945$, January 2021). For this purpose, I recruit new workers and elicit their effort choice for *both* the high and the low piece rate. Workers commit to how many tasks they would complete for both piece

[15] I also replicate the results in an analogous comparison of the "equal rates" control condition with an additional "equal rates" attention condition ($n = 267$, Appendix C.2).

[16] Here, the line between cognition-based and preference-based explanations becomes blurred. Spectators might discount the uncertain counterfactual because doing so is cognitively less demanding or because they prefer to base their merit judgments on hard evidence rather than mere conjectures.

Table 5 Experimental conditions in the counterfactual study

	(1)	(2)	(3)	(4)-(7)
	Actual effort share of worker B			
Effort scenario	0%	10%	30%	Random*
Counterfactual effort share of worker B, by experimental condition				
No information	–	–	–	–
Low counterfactual	0%	10%	30%	Random*
High counterfactual	50%	50%	50%	Random*

*Effort choices: E_A is uniformly randomly drawn from the integers between 0 and 50. E_B ranges from 0 to 25. Counterfactual effort choice of worker B: C_B ranges from E_B to 50.
Notes: This table presents an overview of all seven effort scenarios and the experimental conditions in the counterfactual study. A contingent response method is used: Each spectator faces eight effort scenarios. The seven scenarios above are hypothetical. An eighth effort scenario (not shown) is real. Spectators do not know which scenario is real and have to take each of their decisions seriously. Scenarios (1) to (3) provide the reduced-form evidence analyzed in this section. They are presented in random order to spectators. Data from scenarios (4) to (7) are used in Section 5.4 to structurally estimate a model of merit views.

rates, are then randomly assigned to one piece rate, and subsequently have to follow-up on their commitment. Importantly, this technique measures workers' counterfactual effort choice in an incentivized way. Thus, I know how many tasks the workers (would) complete for both piece rates. Spectators are informed about this procedure.

The spectator sample broadly represents the US population (Appendix Table A.2). As before, spectators make merit judgments in eight scenarios of which the first seven are hypothetical (contingent response method). Spectators do not know which of the eight scenarios is real, so all their decisions are probabilistically incentivized. The first three scenarios are taken from the main experiment and are presented in random order. Here, the advantaged worker A completes more tasks than the disadvantaged worker B, that is, 50 to 0 tasks ($e = 0\%$), 45 to 5 tasks ($e = 10\%$), or 35 to 15 tasks ($e = 30\%$).[17] The next four scenarios are randomly generated and will be used for robustness analyses and in the structural estimation of Section 5.4.

Spectators are randomized into one of three experimental conditions. The conditions vary whether and what spectators learn about what the disadvantaged worker would have done in the advantaged environment. Table 5 provides an overview of all effort scenarios and experimental conditions. Treatment assignment is balanced across covariates (Appendix Table A.5).

No information condition (short: None): No information about worker B's counterfactual effort choice is provided. Therefore, the condition replicates the main treatment condition and serves as a baseline condition in this experiment.

[17] In the other scenarios of the main experiment, the disadvantaged worker completes the same or a larger number of tasks than the advantaged worker. These scenarios are not compatible with the "high counterfactual" condition and are therefore not included.

Low counterfactual condition (short: Low): Spectators are informed about the counterfactual effort choice of worker B for a high piece rate. Worker B would not change his effort provision and thus would not exert more effort for a higher piece rate. This also means that worker B's effort choice is not shaped by his circumstances.

High counterfactual condition (short: High): This condition also provides information about the counterfactual effort choice of worker B. Here, however, worker B would complete as many tasks as worker A for a high piece rate. Thus, the low piece rate of worker B strongly shapes his effort choice. In fact, worker A and worker B (would) make the same choices in the advantaged environment.

Results Figure 3 presents the results, using the same format as earlier figures (see also Appendix Table C.3). First, it reveals that the average reward for worker B is very similar in the "no information" condition and the "low counterfactual" condition. If at all, spectators are even slightly more generous towards worker B in the "low counterfactual" condition. Thus, workers with unknown counterfactual are not treated better than workers whose counterfactual is verifiably low. This confirms that spectators in the baseline condition make their merit judgments as if choices had not been shaped by circumstances. In the presence of an unknown, uncertain counterfactual, they base their merit judgments on the only clear and "hard" evidence they have, namely observed effort choices.

Second, a comparison of the conditions "low counterfactual" and "high counterfactual" exposes that, once known, the counterfactual choice of worker B substantially matters for the merit judgments of spectators. Spectators distribute on average a 9.7 pp higher payment share to worker B when they know that he would have worked as hard as worker A, had he earned a high piece rate (Panel A of Figure 3).[18] This almost doubles the reward share that the disadvantaged worker receives across scenarios. Moreover, the effect occurs in all three effort scenarios (Panel B). For example, if worker A completes 50 tasks while worker B is not willing to work at all, worker B receives a 12 pp higher reward share if spectators know that he would also have completed 50 tasks for a high piece rate.

In light of the framework discussed in Section 2, the evidence implies that comparable choice meritocrats exist but do not apply their merit view when the counterfactual effort choice under equal circumstances is uncertain and unknown. Consequently, dis-

[18]Could the large effect of the "high counterfactual" treatment be partially driven by an experimenter demand effect? The null result in the attention experiment renders such an explanation unlikely. Here, the scope for demand effects appears to be greater. Respondents receive two pages of information that strongly emphasize that choices are shaped by circumstances. Nonetheless, I do not find a treatment effect, suggesting that demand effects are not an empirically important factor in the experimental context of this study.

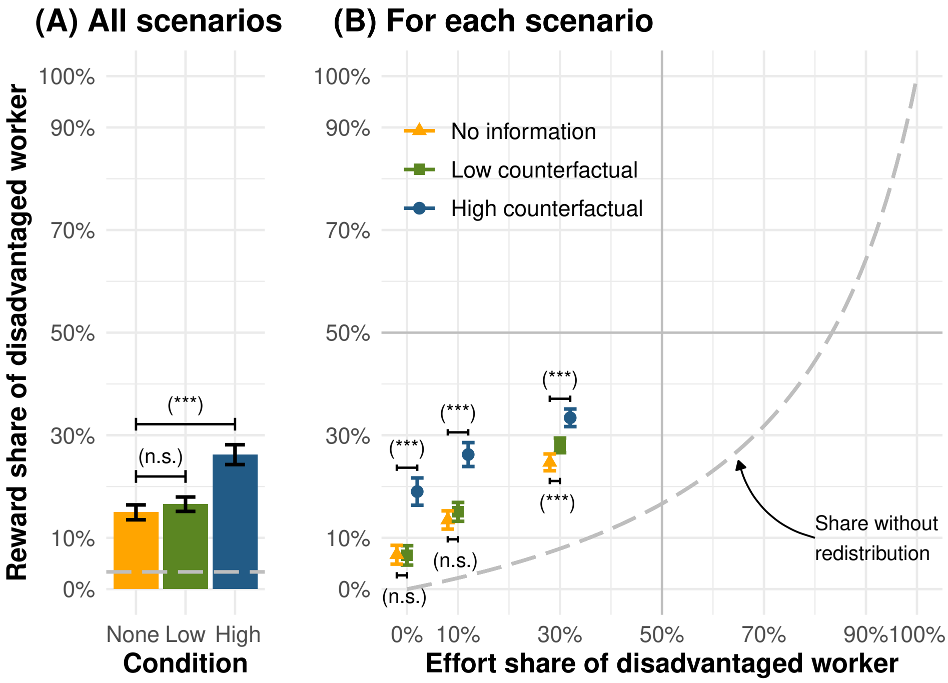


Figure 3 **Counterfactual study**: Mean reward share of disadv. worker (95% CI)

Notes: Results from the counterfactual study, decisions 1-3. Panel A displays the mean reward share assigned to the disadvantaged worker B in each experimental condition, averaged across all three effort scenarios, with 95% confidence intervals. Panel B plots the mean reward share in each effort scenario with 95% confidence intervals. The gray dashed line shows the default share, that is, which payment share worker B would receive if spectators do not redistribute. I test for differences between the "High counterfactual" and the "No information" condition (upper test) and between the "Low counterfactual" and the "No information" condition (lower test). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, (n.s.) $p \geq 0.10$.

advantaged workers do not face a benefit but rather a "burden of the doubt": They cannot verify what they would have done under better circumstances and are hence judged by their actual choices, even though these choices are disadvantageously shaped by unequal circumstances.

Result 3: Once the uncertainty of the counterfactual state is resolved, spectators compensate workers whose choice is shaped by disadvantageous circumstances.

Robustness I replicate the results in two robustness checks. First, the evidence presented above contrasts two extreme counterfactual cases: the disadvantaged worker reacts either strongly or not at all to the piece rate. To test whether merit judgments also respond to counterfactual choices in intermediate cases, I use the data from scenarios 4–7 that randomly vary the actual effort share and the counterfactual effort share of the disadvantaged worker (see Table 5). Do spectators' merit judgments respond to small changes in counterfactual choices? Appendix Figure C.3 confirms this. Spectators assign monotonically higher reward shares for higher counterfactual effort shares.

Second, the counterfactual study provides information on what the disadvantaged worker would have done in the advantaged environment. In a robustness study ("advantaged counterfactual study", $n = 893$, June 2022), I field an analogous experiment with one critical difference: the study provides information on what the *advantaged worker* would have done in the *disadvantaged environment*. This counterfactual allows spectators to compare the effort of both workers under identically disadvantaged circumstances. The robustness experiment thus provides an alternative independent test for the relevance of comparable counterfactuals. Reassuringly, I obtain qualitatively and quantitatively very similar results (see Appendix C.4).

5.4 Heterogeneous fairness views

An analysis of average rewards is blind to the heterogeneity of spectators' merit judgments, but this variation is critical to understand why spectators neglect that circumstances shape choices. Are all spectators sensitive to workers' counterfactual choices, or do some spectators continue to ignore that circumstances shape choices even when the counterfactual state is known? This subsection sheds light on the heterogeneity in spectators' merit judgments *when the counterfactual state is known*. I proceed in three steps. First, I exemplify that the merit judgments of spectators are substantially heterogeneous. Second, I use open-ended text data to confirm that this variation reflects fundamentally different fairness views. Third, I assess the prevalence of these different fairness views with the help of a simple structural model. I conclude that only a subset of spectators are responsive to counterfactual states. Many spectators are actual choice meritocrats who deliberately hold workers responsible for their effort choices even if these choices result from unequal circumstances.

Heterogeneity in merit judgments Figure 4 displays a histogram of the reward decisions made in the counterfactual experiment. For simplicity, I focus on the third effort scenario of the "high counterfactual" condition. In this scenario, worker A completes 35 tasks, while worker B completes only 15 tasks, but it is known that worker B would also have been willing to complete 35 tasks for the high piece rate. The distribution of merit judgments is substantially heterogeneous. It exhibits three discrete spikes. Most pronounced are the spikes around the reward shares of 30%, equal to the actual effort share of worker B, and 50%, equal to the counterfactual effort share of worker B. A third spike is visible around 8%, equal to the initial payment share of worker B before redistribution. 87% of all merit judgments are in the immediate neighborhood (± 5 pp) of these spikes. The data thus clearly show that spectators' merit judgments fall into different types. These types are stable across scenarios. For example, among the spectators who reward workers according to their actual effort in scenario 3, 83% and 73% follow the same approach in scenarios 1 and 2. Likewise, among the spectators who split the rewards equally in scenario 3, 71% and 83% follow the same approach in the

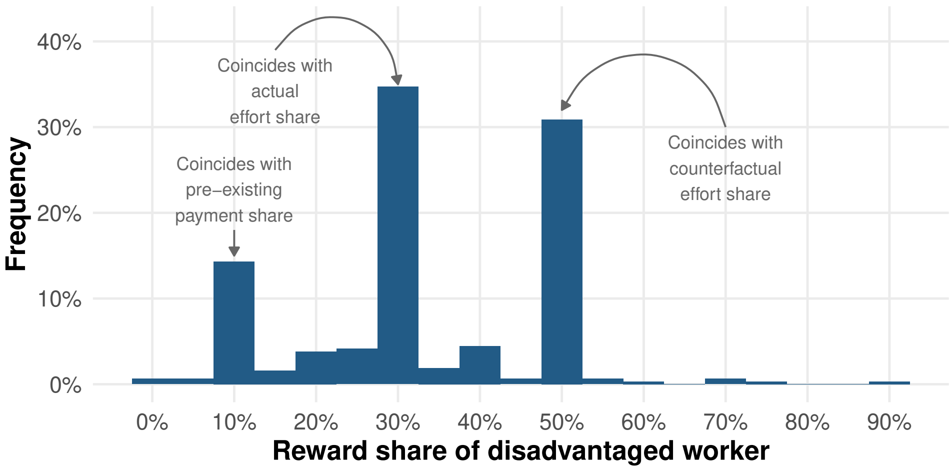


Figure 4 Histogram of merit judgments in the scenario with an actual effort share of 30% and a known counterfactual effort share of 50%

Notes: Results from the counterfactual study, scenario 3, condition: "high counterfactual" (n = 314). Histogram of the reward share that spectators assign to the disadvantaged worker.

scenarios 1 and 2.

Heterogeneous merit judgments reflect different fairness views These different reward patterns coincide with distinct fairness views. For example, spectators who give a 50% share to the disadvantaged worker might act in accordance with the fairness ideal of comparable choice meritocratism. Recall that comparable choice meritocrats think that the disadvantaged worker B deserves a payment share equal to the counterfactual effort share of 50% that he would have provided had he been in the same advantaged circumstances as worker A. On the other hand, spectators who give a 30% reward share to the disadvantaged worker might act in accordance with the fairness ideal of actual choice meritocratism. Actual choice meritocrats hold that the disadvantaged worker B deserves a payment share equal to his effort share of 30%, irrespective of whether effort choices are shaped by external circumstances.

But are these fairness ideals also on top of spectators' minds? To find out, I run an additional study (rationale study, n = 197, September 2022, Prolific US, Appendix C.5) in which I ask spectators for the rationale behind their reward decision in the scenario studied above (A completes 35 tasks; B completes 15 tasks, but would have completed 35 tasks for the high piece rate).[19] Spectators can describe their reasoning in an open-text box which allows them to articulate their thoughts without any restrictions imposed by the researcher. Their responses show that concerns about fairness are ubiq-

[19] I focus on one scenario and one redistribution decision to facilitate the open-ended explanation. However, it implies that I cannot employ an incentivized contingent response method, which is why spectators' redistribution decisions are hypothetical. The average reward assigned to the disadvantaged worker (32%) is quantitatively very close to the average reward share observed in the incentivized counterfactual study (33%).

Table 6 Examples of spectators' rationales

Comparable choice meritocrats	**Actual choice meritocrats**
"Worker B contributed less, but that was due solely to their random assignment to the lower tier of payment. Since that was outside of worker B's control, and they indicated that would have done more if paid more, and worker A's higher pay and performance was also due to chance, I decided to split the proceeds equally."	"I believe that people should be compensated proportionally for the work they completed. If worker B did 30% of the work, he should get 30% of the reward."
	"It seemed fair to me that the pay was split the same way the amount of work was. While I know worker B committed to doing 50% of the work had they been selected [for the high piece rate], they still weren't, so it seems unfair to take pay away from the worker who did actually complete the work."
"The random method by which these two workers received their piece work rate was very unfair. I would not blame worker B for doing less tasks when B was receiving such a paltry rate for their work. So even though A did more work, I felt it was only fair they should be compensated equally. B indicated that B was willingly to do more tasks if B had been compensated more fairly."	"If worker B wasn't doing as much as worker A, why should they be equally compensated? I understand worker B could've be selected to do that much work, but as it stands, I went with what was. Not what could've been."

Notes: Example responses from the rationale study.

uitous. Virtually all responses make a case for which rewards are "fair", which rewards workers "deserve", or "should" receive. 71% of the spectators even explicitly use one or more of these words in their explanation. I manually classify the responses into the fairness views of actual choice meritocrats, comparable choice meritocrats, or a residual category (see C.5) and find that 73% clearly refer to one of the two meritocratic fairness ideals.[20,21] The examples in Table 6 illustrate that their rationales can be quite sophisticated. Taken together, the open-text data thus corroborate that the merit judgments of spectators reflect different fundamental views on fairness.

The prevalence of different fairness views How prevalent are these different fairness views in the full US population? This question can best be assessed if data from many different scenarios are considered jointly. Therefore, I use the merit judgments made in the randomly generated scenarios with known counterfactual state (scenarios 4–7, counterfactual study, see Table 5). These scenarios randomly vary the actual and counterfactual effort shares of both workers and cover a rich variety of cases. I aggregate and interpret the data with the help of a simple structural model.

[20] 21% of respondents argue like comparable choice meritocrats, and 52% argue like actual choice meritocrats. These figures are in the same ballpark as the estimates from the structural model (see below). The structural estimates have the advantage that they draw on data from many different scenarios and use data from a broadly representative sample.

[21] Their responses are highly predictive of their reward decisions. Spectators whose rationale reflects comparable choice meritocratism are 71 pp more likely to assign a reward share to the disadvantaged worker that is higher than his actual reward share ($p < 0.001$). Spectators whose rationale reflects actual choice meritocratism are 73 pp more likely to distribute the rewards according to the actual effort shares of the workers ($p < 0.001$).

In line with Almås et al. (2020), I assume that spectator i selects a reward share r_i for the disadvantaged worker to maximize the utility function

$$U(r_i) = -\left[r_i - m_i(s)\right]^2$$

where $m_i(s)$ denotes i's merit view, that is, her view of the reward that the disadvantaged worker deserves in situation s. A situation $s = (e, c, p)$ is characterized by the actual effort share of the disadvantaged worker e, his counterfactual effort share c, and the pre-existing reward share p. The spectator wants to implement the reward share r_i that she thinks is merited by worker B: $r_i^* = m_i(s)$. However, the decisions of spectators are noisy and deviate from their merit views by a normally distributed response error $\varepsilon_{is} \sim_{iid} N(0, \sigma^2)$.

$$\hat{r}_i^* = m_i(s) + \varepsilon_{is}$$

The model assumes that the population is divided into four distinct fairness types. *Actual choice meritocrats* and *comparable choice meritocrats* have been introduced before. *Libertarians* regard any pre-existing earning share p as legitimate and thus fully accept the pre-existing inequality. *Egalitarians* hold that the workers always deserve equal payment shares and thus always implement equality. In short,

$$m_i^t(s) = \begin{cases} e & \text{for } t = \text{actual choice meritocrat} \\ c & \text{for } t = \text{comparable choice meritocrat} \\ p & \text{for } t = \text{libertarian} \\ \frac{1}{2} & \text{for } t = \text{egalitarian} \end{cases}$$

I estimate five parameters, namely the population shares of each preference type together with the standard deviation of the response error σ. I employ a constrained maximum likelihood procedure.[22,23]

The model estimates that 40% of the population are actual choice meritocrats, while 28% are comparable choice meritocrats. Libertarians and egalitarians have a population share of 16% and 15%, respectively (see Table 7). The estimates confirm that the large majority of participants, approximately 70%, endorse a meritocratic fairness ideal.[24]

[22]Appendix C.6 presents the technical details of the estimation procedure and shows that the results are robust to a series of sensitivity checks, such as a specification with trembling-hand response error, an exclusion of participants who initially failed a control question, or the introduction of a fifth "noise" type. I also confirm the numerical stability of the maximum likelihood estimator in Monte Carlo experiments.

[23]An earlier version of paper, archived at https://osf.io/7tkpe/, presented a more general version of the structural model. Run with data from all conditions of the counterfactual study, the model also estimates the discount factor with which comparable choice meritocrats discount uncertain counterfactual states. The results confirm that they fully discount uncertain counterfactual states, replicating the reduced-form results from Section 5.3.

[24]The estimated share of meritocrats is much higher than in Almås et al. (2020), who classify 37.5%

Table 7 Results of the structural estimation

	Estimate	95% confidence interval
Population shares		
Actual choice meritocrats	40.0%	[35.9% – 44%]
Comparable choice meritocrats	28.4%	[24.7% – 32.2%]
Libertarians	16.2%	[13.3% – 19.2%]
Egalitarians	15.4%	–
Error term and sample		
σ noise	9.58	[9.30 – 9.85]
Respondents	630	
Decisions	2520	

Notes: Results from the counterfactual study, decisions 4–7, maximum likelihood estimation of the structural model of fairness views. The estimates indicate the population shares of different fairness views. No confidence interval is reported for the share of egalitarians because their share is deduced from the other estimates. See Appendix C.6 for further details.

However, they also confirm that many meritocrats are actual choice meritocrats. For them, it is irrelevant that workers' choices are shaped by unequal circumstances, even if they know what would have happened under equal circumstances. I do not detect statistically reliable differences in the composition of fairness types across demographic groups.[25]

> **Result 4**: Merit judgments are substantially heterogeneous and reflect different fundamental fairness views. A structural model of merit views classifies only 28% of individuals as comparable choice meritocrats who want to correct for the effect of circumstances on choices. 40% of individuals are actual choice meritocrats who assign rewards based on effort, even when effort choices have verifiably been shaped by unequal circumstances.

Therefore, the insensitivity of merit judgments to the fact that circumstances shape choices derives from two complementary sources: For some spectators, it derives from actual choice meritocratism, their personal view on fair rewards. For others, it results from the uncertainty of the counterfactual. Even though they are comparable choice meritocrats, they do not factor in the effect of unequal circumstances on choices if it remains uncertain what exactly would have happened on a level playing field.

of the US population as meritocrats. However, in their setting, spectators receive only coarse, binary information about effort choices, namely which of two workers is more productive. Merit plays a much stronger role in my setting because spectators learn exactly how many tasks each worker completed.

[25] To test whether the composition of fairness types varies across different parts of the population, I re-estimate the model and allow its parameters to vary across two separate groups of the population (see Appendix C.6). In separate analyses, I compare female versus male respondents, respondents with above- versus below-median income, respondents with versus without a college degree, and Republicans versus Democrats. I find that female spectators are 8 pp more likely to be actual choice meritocrats, and high-income and college-educated spectators are 6 pp less likely to be libertarians. However, all three effects are only marginally significant (see Appendix Table C.5). All other patterns are difficult to interpret, given the lack of statistical significance.

6 Merit judgments in real-world scenarios

The controlled environment of the choice experiment has critical advantages. In particular, it measures merit judgments in situations with real consequences, it exogenously varies whether circumstances differentially shape choices, and it allows for a detailed exploration of the underlying mechanism. However, the stylized environment also comes at a cost: It differs from many real-life settings that characterize the debate about meritocracy.

In this section, I therefore explore whether merit judgments are also insensitive to the effect of circumstances on choices in three real-world scenarios. I report results from an additional vignette study which sheds light on the following three questions, chosen as common and important practical examples of merit judgments. First, revisiting the example of racial discrimination in the labor market discussed in Section 2, are minorities compensated for the detrimental choices they might make because they are discriminated? Second, is a person growing up with few opportunities and incentives to exert effort blamed for being idle? And third, is an entrepreneur rewarded for taking the risk of founding a company if he inherited a fortune so generous that it made founding easy and substantially reduced any risk involved?

6.1 Design

The vignette study was conducted in February 2021 in collaboration with the survey company Lucid. Respondents were recruited from the general US population (n = 1,222).[26] Each vignette describes a simple hypothetical scenario with two persons who are exposed to unequal circumstances. Influenced by these circumstances, the disadvantaged person makes a detrimental choice and, as a consequence, earns much less money than the other person. Below, I outline each vignette. The complete instructions of the vignette study are available online (https://osf.io/xj7vc/).

Discrimination vignette: A white and a black employee compete for a promotion. However, their boss is notorious for being racist and has never promoted

[26]The study was conducted in two waves. Wave 1 was collected together with the disappointment study. Here, each respondent faced two randomly selected vignettes. Wave 2 was launched shortly thereafter, and respondents faced all vignettes in random order. Respondents who speed through the survey and complete vignettes with an average response time of less than one minute are excluded. The results are robust to both stricter and more lenient exclusion criteria (see Table D.1). Table A.2 shows that the sample does not fully match the characteristics of the general population. Among others, the sample contains more women, more older respondents, and more respondents with a low income. However, the results are robust to the use of survey weights that correct for these imbalances (see Table D.1). The vignette survey also contained a fourth vignette on criminal behavior which requires a tailored analysis and discussion and is not reported here for brevity (but see the earlier version of the paper, archived at https://osf.io/7tkpe/).

employees of color before. The white employee decides to work hard to win the promotion, the black person does not. In the end, the white employee is promoted and receives an attractive one-time bonus of $10,000.

Poverty vignette: In this vignette, the advantaged person grew up in a rich family, went to good schools, and was taught that "you can go as far as your hard work takes you." The disadvantaged person grew up in a poor family, went to poor-quality schools, and was always told that "the poor stay poor, and the rich get richer." Whereas the advantaged person always worked hard in his life and, as a consequence, earns $125,000 a year, the disadvantaged person never worked hard and earns only $25,000 a year.

Entrepreneur vignette: The entrepreneur vignette portrays two passionate software developers who always dreamed of founding a software start-up. The advantaged person inherited a considerable fortune that provided him with enough money to found and fail several times without any risk of financial ruin. By contrast, the disadvantaged person would have struggled to gather enough money to launch even a first start-up and would have been broke if his first attempt had failed. The advantaged person decided to take the risk and founded his own software start-up. He earns $200,000 a year today. The disadvantaged person decided to work as a software developer for a local company. He earns $50,000 a year today.

Similarly to the main experiment, respondents can specify how much money each person deserves by hypothetically redistributing the income (or bonus) between the two people. If their merit judgments are sensitive to the effect of circumstances on choices, they should compensate the disadvantaged person. However, redistribution towards the disadvantaged person could also be explained by other fairness motives. In particular, respondents might assign more money to the disadvantaged person simply because they prefer a more equal outcome. Or they want to compensate the disadvantaged person for living in worse circumstances, for example, for not inheriting any money in the entrepreneur vignette. To identify the sensitivity of merit judgments to the effect of circumstances on choices, I introduce a between-subject variation that is analogous to the counterfactual study of Section 5.3. Respondents are randomized into one of three treatments. The treatments vary whether and what spectators learn about what the disadvantaged person would have done in the advantaged environment.

Baseline condition: The vignettes describe only the actual decisions of both persons.

Low counterfactual condition: Each vignette states that the disadvantaged person would not have made a different choice if he had been in the advantaged situation. Hence, his choice was not shaped by his circumstances.

High counterfactual condition: Here, the disadvantaged person would have made the same choice as the advantaged person if he had been in the advantaged situation. Therefore, his choice was strongly shaped by his circumstances.

6.2 Results

Table 8 summarizes the results. Once again, I find that merit judgments are insensitive to the effect of circumstances on choices. First, I observe only little redistribution towards the disadvantaged person in the baseline condition. For example, in the discrimination vignette, only 42% of the respondents assign a positive reward share to the discriminated black employee (Column 1, Panel A), and, on average, he receives only 14% of the total payoff (Column 1, Panel B). Most respondents accept that he comes away empty-handed. His choice not to work hard legitimizes the highly unequal outcome. In the poverty vignette, 55% of the respondents are willing to compensate the person who grew up in poverty, but he is still assigned only 24% of the total earnings (only 7 pp more than he would receive without redistribution).

Next, I study the difference in merit judgments between the baseline and the "low counterfactual" condition. In the baseline condition, circumstances shape the actors' choices (though the counterfactual is uncertain), whereas choices are verifiably unaffected by circumstances in the "low counterfactual" condition. If, as in the main experiment, baseline merit judgments are insensitive to the effect of circumstances on choices, they should not vary between the baseline and the "low counterfactual" condition. Indeed, I find that the reward decisions are similar in both conditions. Pooled across vignettes, only 0.4 pp more respondents redistribute money to the disadvantaged person in baseline than in "low counterfactual" (Column 4, Panel A). Likewise, the average reward share of the disadvantaged person is only 1.5 pp higher in the baseline condition (Column 4, Panel B). Both effects are statistically insignificant.

In stark contrast, the "high counterfactual" condition increases the share of respondents who redistribute money towards the disadvantaged person by 12.6 pp and raises his mean reward share by 6.8 pp across vignettes. The results are mainly driven by the discrimination and the poverty vignette. For instance, in the discrimination vignette, 23 pp more respondents are willing to assign a positive reward share to the black employee once they know that he would have worked equally hard had his boss given him a fair chance. Likewise, the fraction of respondents who compensate the disadvantaged person increases by 9 pp in the poverty vignette. Respondents thus integrate circumstances' effect on choices in their merit judgments when the counterfactual is known, but few do so if the counterfactual is uncertain. The effect is more muted in the

Table 8 Merit judgments in the vignette study

(A) Share of respondents redistributing towards the disadvantaged worker

	Binary indicator for compensation			
	Discrimination	Poverty	Entrepreneur	Pooled
	(1)	(2)	(3)	(4)
Low counterfactual	0.015	-0.001	-0.026	-0.004
	(0.041)	(0.041)	(0.040)	(0.029)
High counterfactual	0.230***	0.090**	0.059	0.126***
	(0.040)	(0.040)	(0.039)	(0.029)
Constant	0.424***	0.547***	0.630***	
	(0.028)	(0.028)	(0.028)	
Vignette FE	–	–	–	✓
Observations	889	887	888	2,664
R^2	0.044	0.008	0.005	0.587

(B) Mean reward share of disadvantaged person

	Reward share of disadv. person (in %)			
	Discrimination	Poverty	Entrepreneur	Pooled
	(1)	(2)	(3)	(4)
Low counterfactual	0.133	-2.387*	-2.391	-1.539
	(1.658)	(1.197)	(1.413)	(1.085)
High counterfactual	13.590***	4.003***	2.867*	6.795***
	(1.797)	(1.277)	(1.463)	(1.177)
Constant	13.994***	24.208***	33.497***	
	(1.182)	(0.874)	(1.044)	
Initial reward share	0.00	17.00	20.00	
Vignette FE	–	–	–	✓
Observations	889	887	888	2,664
R^2	0.082	0.029	0.015	0.683

Notes: Results from the vignette study, OLS regressions, robust standards (Columns 1-3) and standard errors clustered at the respondent level (Column 4) in parentheses. The dependent variable in Panel A is a binary indicator for whether a respondent compensates the disadvantaged person by redistributing money towards him. The dependent variable in Panel B is the reward share assigned to the disadvantaged person. The independent variables are treatment dummies. Columns 1-3 report results from different vignettes, and Column 4 displays the pooled results. In each panel, p-values of the coefficients in Columns 1-3 are adjusted for multiple hypothesis, using the Benjamini-Hochberg adjustment. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

entrepreneur vignette where the merit judgments of respondents appear to be largely insensitive even to information about counterfactual states.

Finally, I also investigate the insensitivity to the effect of circumstances on choices through the lens of qualitative survey questions. Do respondents describe it as "fair" that agents' unequal choices result in unequal outcomes, even though they consider it "unfair" that two agents faced unequal circumstances? To test this, I run a short, complementary survey ("vignette evaluation" study, $n = 601$, September 2022, Prolific US). Respondents face one randomly chosen vignette and select whether they evaluate it as "fair" or "unfair" that (i) the two persons in a vignette experienced unequal circumstances and (ii) unequal outcomes result from their different choices. I find that, even though most of the respondents view the unequal circumstances in the three vignettes as unfair (discrimination: 81%, poverty: 73%, entrepreneur: 74%), many consider the unequal outcomes of the vignettes as fair (discrimination: 96%, poverty: 82%, entrepreneur: 67%). Moreover, many respondents express both views simultaneously. Among respondents who view unequal circumstances as unfair, large majorities evaluate the outcomes of the vignettes as fair (discrimination: 95%, poverty: 75%, entrepreneur: 55%).

Taken together, the results suggest that merit judgments are insensitive to the effect of circumstances on choices not only in the controlled experimental setting but that the same phenomenon is to be expected in many important real-world domains of a meritocracy.

Result 5: Merit judgments do not factor in the effect of circumstances on choices in important real-world scenarios.

7 Conclusion

The idea of meritocracy has become central in Western politics, where it has shaped the public debate, the political and economic culture, and social reforms. Meritocracy promises that the family, neighborhood, and other circumstances into which one is born should not matter. This promise is popular and closely related to the prominent ideas of equal opportunity and the American dream.

However, the results of a series of experiments with more than 6,000 US participants suggest that the prevailing notion of meritocratic fairness is *shallow*. Circumstances often shape the choices that agents make, and people's merit judgments tend to be insensitive to this effect. People hold others responsible for their choices, even when these choices are strongly shaped by external circumstances. Evidence on the mechanism behind this phenomenon suggests that it is likely to be a fundamental feature of merit judgments. First, about one-quarter of participants would, in principle, prefer to compensate agents for the disadvantageous effects of circumstances on choice. Yet, they abstain from doing so unless they verifiably know what would have happened on a level playing field. Such certainty about the counterfactual state is extremely rare in the real world and will thus often form a binding constraint for merit judgments. Second, many individuals do not factor in the effect of circumstances on choices, even when they are fully informed about the counterfactual.

These results refine our understanding of the popular notion of meritocratic fairness and have important implications for the debate about equal opportunity. First, in a shallow meritocracy, the disadvantaged can be doubly disadvantaged. When unequal circumstances impede the accumulation of merit because they discourage hard work or stifle ambitions, disadvantaged agents not only face adverse and discouraging circumstances, but they are also blamed and regarded as undeserving if they show less dedication and perseverance in these circumstances. Their choices and achievements are measured with the same yardstick as those of advantaged groups, even though their starting position is different. In the terminology of Bohren et al. (2022), this is a case of systemic discrimination. Second, affirmative action policies, which aim to correct for unequal opportunities that agents face in producing merit, undermine the prevailing notion of meritocratic fairness. This could explain why they belong to the most controversial policy issues (Harrison et al., 2006). Third, for shallow meritocrats, predistributive and redistributive policies differ in a critical respect: Predistribution equates circumstances ex ante. It thus prevents that a differential effect of circumstances on choices occurs, and shallow meritocrats will endorse the accompanying increase in equal opportunities. By contrast, redistribution, even if it is targeted to compensate for unequal opportunities, intervenes only after unequal circumstances led to unequal choices. Insofar as this clashes with the principle of responsibility for choices, it is likely to meet resistance among shallow meritocrats.

In light of these consequences, an important avenue for future research is to identify when and how unequal socioeconomic circumstances shape important life choices, such as working hard, taking risks, or having ambitious career aspirations (Altmejd et al., 2021; Bursztyn et al., 2017; Carlana et al., 2022; Glover et al., 2017). Such research will reveal the contexts in which shallow meritocracy matters most and where merit judgments are susceptible to ignore sizable effects of circumstances on choices. In addition, not only choices, but also valued abilities such as cognitive skills are considered important determinants of merit. Future research is needed to explore whether people's evaluation of skills is similarly blind to the circumstances that foster or impede their development. Likewise, it would be fruitful to investigate how common the neglect of circumstances' effect on choices is outside the US, in countries with different cultures or welfare state regimes.

The pros and cons of meritocracy have been subject to a heated public debate (Frank, 2016; Greenfield, 2011; Markovits, 2019; Sandel, 2020; Wooldridge, 2021; Young, 1958). In view of this debate, it seems warranted to ask to what extent we should actually be concerned about meritocracy being shallow. This is a normative question open for discussion, but I briefly sketch two possible perspectives. On the one hand, one could think of shallow meritocracy as a problematic flaw in merit judgments. It arguably appears inconsistent to acquit people from their circumstances but at the same time hold them responsible for the choices that these circumstances promote and produce. On the other hand, this behavior might constitute a second-best response to a world of limited information. After all, neither the ultimate cause of each decision nor the decisions that would have been made in counterfactual states of the world are known. Holding others responsible for their choices could be an adaptive, simple shortcut, a societal rule-of-thumb. It provides clear incentives to agents and clear guidance to spectators. Ultimately, shallow meritocracy and responsibility for one's choices may simply be a practical necessity of living together. In either case, those opposed to shallow meritocracy have a strong argument for advancing equal opportunities. Equal opportunities level the effect of circumstances on choices and thus also defuse the problem of shallow meritocracy.

References

Abadie, Alberto, "Statistical Nonsignificance in Empirical Economics," *American Economic Review: Insights*, 2020, *2* (2), 193–208.

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

Akerlof, George A. and Janet L. Yellen, "The Fair Wage-Effort Hypothesis and Unemployment," *The Quarterly Journal of Economics*, 1990, *105* (2), 255–283.

Alan, Sule and Seda Ertac, "Fostering Patience in the Classroom: Results from Randomized Educational Intervention," *Journal of Political Economy*, 2018, *126* (5), 1865–1911.

Alesina, Alberto and Edward Glaeser, *Fighting Poverty in the US and Europe: A World of Difference*, Oxford University Press, 2004.

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

— **, Stefanie Stantcheva, and Edoardo Teso**, "Intergenerational Mobility and Preferences for Redistribution," *American Economic Review*, 2018, *108* (2), 521–554.

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

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith, "O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries," *The Quarterly Journal of Economics*, 2021, *136* (3), 1831–1886.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples," *Review of Economic Studies*, 2022, *rdac008.*

Andreoni, James, Deniz Aydin, Blake Allen Barton, B. Douglas Bernheim, and Jeffrey Naecker, "When Fair Isn't Fair: Understanding Choice Reversals Involving Social Preferences," *Journal of Political Economy*, 2020, *128* (5), 1673–1711.

Baron, Jonathan and John C. Hershey, "Outcome Bias in Decision Evaluation," *Journal of Personality and Social Psychology*, 1988, *54* (4), 569–579.

Bartling, Björn, Alexander W. Cappelen, Mathias Ekström, Erik Ø. Sørensen, and Bertil Tungodden, "Fairness in Winner-Take-All Markets," *Working Paper*, 2018.

— **and Urs Fischbacher**, "Shifting the Blame: On Delegation and Responsibility," *The Review of Economic Studies*, 2012, *79* (1), 67–87.

Benjamin, Daniel J., "Errors in probabilistic reasoning and judgmental biases," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics: Applications and Foundations 2*, North-Holland, 2019, chapter 2.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir, "A Behavioral-Economics View of Poverty," *American Economics Reveiw*, 2004, *94* (2), 419–423.

Bohren, J. Aislinn, Peter Hull, and Alex Imas, "Systemic Discrimination: Theory and Measurement," *Working Paper*, 2022.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini, "Identity, Beliefs, and Political Conflict," *The Quarterly Journal of Economics*, 2021, *136* (4), 2371–2411.

Brandts, Jordi and Gary Charness, "The strategy versus the direct-response method: a first survey of experimental comparisons," *Experimental Economics*, 2011, *14*, 375–398.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Brownback, Andy and Michael A. Kuhn, "Understanding outcome bias," *Games and Economic Behavior*, 2019, *117*, 342–360.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, *107* (11), 3288–3319.

Bushong, Benjamin and Tristan Gagnon-Bartsch, "Reference Dependence and Attribution Bias: Evidence from Real-Effort Experiments," *American Economic Journal: Microeconomics*, 2022.

Byrne, Ruth M.J., "Counterfactual Thought," *Annual Review of Psychology*, 2016, *67*, 135–157.

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

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

_ , James Konow, Erik Ø. Sørensen, and Bertil Tungodden, "Just Luck: An Experimental Study of Risk-Taking and Fairness," *American Economic Review*, 2013, *103* (4), 1398–1413.

_ , Karl Ove Moene, Siv-Elisabeth Skjelbred, and Bertil Tungodden, "The merit primacy effect," *Working Paper*, 2020.

_ , Ranveig Falch, and Bertil Tungodden, "Fair and Unfair Income Inequality," in K. F. Zimmermann, ed., *Handbook of Labor, Human Resources and Population Economics*, Springer, 2020, pp. 1–25.

_ , Sebastian Fest, Erik Ø. Sørensen, and Bertil Tungodden, "Choice and Personal Responsibility: What is a Morally Relevant Choice?," *Review of Economics and Statistics*, 2022.

Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti, "Goals and Gaps: Educational Careers of Immigrant Children," *Econometrica*, 2022, *90* (1), 1–29.

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Börn Wallace, "Genetic Variation in Preferences for Giving and Risk Taking," *The Quarterly Journal of Economics*, 2009, *124* (2), 809–842.

Charness, Gary, Uri Gneezy, and Brianna Halladay, "Experimental methods: Pay one or pay all," *Journal of Economic Behavior & Organization*, 2016, *131*, 141–150.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

Coppock, Alexander and Oliver A. McClellan, "Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents," *Research and Politics*, 2019, *6* (1), 1–14.

Davis, Mark H., "Measuring individual differences in empathy: Evidence for a multidimensional approach.," *Journal of Personality and Social Psychology*, 1983, *44* (1), 113–126.

Ding, Peng, Avi Feller, and Luke Miratrix, "Randomization inference for treatment effect variation," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 2016, *78* (3), 655–671.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, "The Intergenerational Transmission of Risk and Trust Attitudes," *The Review of Economic Studies*, 2012, *79* (2), 645–677.

Engl, Florian, "A Theory of Causal Responsibility Attribution," *Working Paper*, 2018.

Enke, Benjamin and Florian Zimmermann, "Correlation Neglect in Belief Formation," *The Review of Economic Studies*, 2017, *86* (1), 313–332.

Esponda, Ignacio and Emanuel Vespa, "Hypothetical Thinking and Information Extraction in the Laboratory," *American Economic Journal: Microeconomics*, 2014, *6* (4), 180–202.

__ **and** __ , "Contingent Preferences and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory," *Working Paper*, 2019.

Falk, Armin, Fabian Kosse, and Pia Pinger, "Mentoring and Schooling Decisions: Causal Evidence," *Working Paper*, 2020.

__ **, Sven Heuser, and David Huffman**, "Moral Luck: Existence, Mechanisms, and Prevalence," *Working Paper*, 2021.

__ **, Thomas Neuber, and Nora Szech**, "Diffusion of Being Pivotal and Immoral Outcomes," *The Review of Economic Studies*, 2020, *87* (5), 2205–2229.

Fisman, Raymond, Ilyana Kuziemko, and Silvia Vannutelli, "Distributional Preferences in Larger Groups: Keeping Up With the Joneses and Keeping Track of the Tails," *Journal of the European Economic Association*, 2020, *jvaa033.*

Frank, Robert H., *Success and Luck: Good Fortune and the Myth of Meritocracy*, Princeton and Oxford: Princeton University Press, 2016.

Gabaix, Xavier, "Behavioral inattention," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics: Applications and Foundations*, Vol. 2, North-Holland, 2019, pp. 261–343.

Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty, "Brahmin Left Versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948-2020," *The Quarterly Journal of Economics*, 2021, *qjab036.*

Giuliano, Paola and Antonio Spilimbergo, "Growing up in a Recession," *The Review of Economic Studies*, 2013, *81* (2), 787–817.

Glover, Dylan, Amanda Pallais, and William Pariente, "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," *The Quarterly Journal of Economics*, 2017, *132* (3), 1219–1260.

Graeber, Thomas, "Inattentive Inference," *Working Paper*, 2022.

Greenfield, Kent, *The Myth of Choice: Personal Responsibility in a World of Limits*, New Haven and London: Yale University Press, 2011.

Gurdal, Mehmet Y., Joshua B. Miller, and Aldo Rustichini, "Why Blame?," *Journal of Political Economy*, 2013, *121* (6), 1205–1247.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," *Journal of Economic Literature*, 2022.

Han, Yi, Yiming Liu, and George Loewenstein, "Confusing Context with Character: Correspondence Bias in Economic Interactions," *Management Science*, 2022.

Harden, Kathryn Paige, *The Genetic Lottery: Why DNA Matters for Social Equality*, Princeton University Press, 2021.

Harrison, David A., David A. Kravitz, David M. Mayer, Lisa M. Leslie, and Dalit Lev-Arey, "Understanding Attitudes Toward Affirmative Action Programs in Employment: Summary and Meta-analysis of 35 Years of Research," *Journal of Applied Psychology*, 2006, *91* (5), 1013–1036.

Haushofer, Johannes and Ernst Fehr, "On the psychology of poverty," *Science*, 2014, *344* (6186), 862–867.

Heckman, James J., "Skill Formation and the Economics of Investing in Disadvantaged Children," *Science*, 2006, *312* (5782), 1900–1902.

Henningsen, Arne and Ott Toomet, "maxLik: A package for maximum likelihood estimation in R," *Computational Statistics*, 2011, *26* (3), 443–458.

Hvidberg, Kristoffer B., Claus T. Kreiner, and Stefanie Stantcheva, "Social Positions and Fairness Views on Inequality," *Working Paper*, 2022.

Kahneman, Daniel and Dale T. Miller, "Norm theory: Comparing reality to its alternatives," *Psychological Review*, 1986, *93* (2), 136–153.

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

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk, "The Formation of Prosociality: Causal Evidence on the Role of Social Environment," *Journal of Political Economy*, 2019, *128* (2), 434–467.

Kovaleva, Anastassiya, *The IE-4: Construction and Validation of a Short Scale for the Assessment of Locus of Control*, Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, 2012.

Krawczyk, Michał, "A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution," *Journal of Public Economics*, 2010, *94* (1-2), 131–141.

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

Lagnado, David A. and Tobias Gerstenberg, "Causation in Legal and Moral Reasoning," in Michael R. Waldmann, ed., *The Oxford Handbook of Causal Reasoning*, New York: Oxford University Press, 2017.

Liang, Yucheng, "Learning from Unknown Information Sources," *Working Paper*, 2021.

Markovits, Daniel, *The Meritocracy Trap*, Penguin Books, 2019.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa, "Failures in Contingent Reasoning: The Role of Uncertainty," *American Economic Review*, 2019, *109* (10), 3437–3474.

Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø. Sørensen, "Luck, choice and responsibility — An experimental study of fairness views," *Journal of Public Economics*, 2015, *131*, 33–40.

Müller, Maximilian W., "Intergenerational Transmission of Education: Internalized Aspirations versus Parent Pressure," *Working Paper*, 2021.

Nagel, Thomas, "Moral Luck," in "Mortal Questions," Cambridge, New York: Cambridge University Press, 1979.

Oprea, Ryan, "Simplicity Equivalents," *Working Paper*, 2022.

Pasek, Josh, Matthew Debell, and Jon A. Krosnick, "Standardizing and Democratizing Survey Weights: The ANES Weighting System and anesrake," *Working Paper*, 2014.

Putnam, Robert D., *Our Kids: The American Dream in Crisis*, New York: Simon and Schuster, 2016.

Roemer, John E., "A Pragmatic Theory of Responsibility for the Egalitarian Planner," *Philosophy & Public Affairs*, 1993, *22* (2), 146–166.

Roese, Neal J., "Counterfactual thinking," *Psychological Bulletin*, 1997, *121* (1), 133–148.

Ross, Lee, "The Intuitive Psychologist and his Shortcomings: Distortions in the Attribution Process," *Advances in Experimental Social Psychology*, 1977, *10*, 173–220.

Sandel, Michael J., *The Tyranny of Merit: What's Become of the Common Good?*, London: Allen Lane, 2020.

Sloman, Steven, *Causal Models: How People Think about the World and Its Alternatives*, New York: Oxford University Press, 2005.

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

Sviatschi, Maria Micaela, "Making a Narco: Childhood Exposure to Illegal Labor Markets and Criminal Life Paths," *Econometrica*, 2022, *90* (4), 1835–1878.

Taubinsky, Dmitry and Alex Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *The Review of Economic Studies*, 2018, *85* (4), 2462–2496.

Wooldridge, Adrian, *The Aristocracy of Talent: How Meritocracy Made the Modern World*, Skyhorse Publishing, 2021.

Young, Michael, *The Rise of the Meritocracy*, Thames and Hudson, 1958.

Online Appendices

Table of Contents

A Samples

Table A.1 Overview of all samples

Sample	When	How	Population	Recruitment	n
Main study					
Main treatment and control condition	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	653
Robustness					
"Equal rates" conditions**	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	661
Disappointment study	February 2021	Online experiment	US adults	Via survey company Lucid	606
Leisure time study	June 2022	Online experiment	US adults	Via survey platform Prolific	1,095
Mechanism					
Attention condition**	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	274
"Equal rates" attention condition**	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	267
Counterfactual study	January 2021	Online experiment	US adults (targeted*)	Via survey company Lucid	945
Advantaged counterfactual study	June 2022	Online experiment	US adults (targeted*)	Via survey company Lucid	893
Rationale study	September 2022	Online survey	US adults	Via survey platform Prolific	197
External validity					
Vignette study	February 2021	Online survey	US adults	Via survey company Lucid	1,222***
Vignette evaluation study	September 2022	Online survey	US adults	Via survey platform Prolific	601
				Total n	6,819

*The sampling process targeted a sample that represents the general population in terms of gender, age (3 groups), region (4 groups), income (3 groups), and education (2 groups). The counterfactual study and the advantaged counterfactual study did not target education.
**Run in parallel to main study.
***Wave 1 of the vignette study was attached to the disappointment study. 595 respondents of the disappointment study also participated in the vignette study. The total n does not double-count these respondents.

Notes: This table provides an overview of all spectator samples used in this study. It lists all experimental conditions and studies and describes when and how they were conducted.

Exclusion criteria in online experiments Exclusion criteria are preregistered (see Appendix F). The samples do not contain the following responses: first, respondents who do not complete the first seven redistribution decisions[27]; second, respondents who spend less than 30 seconds on the instructions until the first treatment variation is introduced; third, duplicate respondents (very rare cases).

Table A.2 Comparison of all samples to the American Community Survey (ACS)

Variable	ACS (2019)	Main study	Equal rates	Disappoint-ment	Leisure time	Attention	Attention equal rates
Gender							
Female	51%	51%	52%	63%	49%	52%	48%
Age							
18-34	30%	30%	28%	11%	47%	32%	33%
35-54	32%	33%	32%	30%	39%	32%	29%
55+	38%	37%	41%	59%	14%	36%	38%
Household net income							
Below 50k	37%	40%	43%	47%	39%	39%	44%
50k-100k	31%	34%	32%	34%	38%	34%	33%
Above 100k	31%	27%	26%	19%	23%	26%	23%
Education							
Bachelor's degree or more	31%	43%	40%	48%	56%	38%	36%
Region							
Northeast	17%	21%	16%	25%	19%	16%	16%
Midwest	21%	21%	22%	25%	20%	18%	21%
South	38%	36%	39%	35%	42%	44%	38%
West	24%	22%	23%	15%	19%	23%	25%
Sample size	2,059,945	653	661	606	1095	274	267

Variable	ACS (2019)	Counter-factual	Adv. counter-factual	Rationale	Vignettes	Vig. eval-uation
Gender						
Female	51%	53%	53%	48%	61%	48%
Age						
18-34	30%	23%	25%	44%	15%	49%
35-54	32%	35%	33%	38%	33%	40%
55+	38%	42%	41%	18%	52%	11%
Household net income						
Below 50k	37%	39%	37%	40%	45%	39%
50k-100k	31%	32%	29%	36%	33%	35%
Above 100k	31%	30%	33%	24%	22%	26%
Education						
Bachelor's degree or more	31%	56%	50%	57%	47%	59%
Region						
Northeast	17%	17%	17%	13%	25%	21%
Midwest	21%	21%	19%	26%	23%	25%
South	38%	38%	40%	38%	36%	43%
West	24%	24%	24%	23%	16%	11%
Sample size	2,059,945	945	893	197	1,222	601

Notes: Column "ACS (2019)" presents data from the American Community Survey (ACS) 2019. The other columns describe the different experimental samples.

[27]There is only one redistribution decision in the disappointment study and the leisure time study. Here, I exclude all respondents who do not complete the study.

Table A.3 Test for balanced treatment assignment – part 1

Main study (treatment vs. control)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	−0.001	0.150	0.754	0.000	−0.012	−0.022	0.031
	(0.039)	(1.339)	(4.488)	(0.039)	(0.032)	(0.038)	(0.032)
Constant	0.511***	47.116***	76.831***	0.426***	0.213***	0.374***	0.204***
	(0.028)	(0.935)	(3.144)	(0.027)	(0.023)	(0.027)	(0.022)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.992$							
Observations	653	653	653	653	653	653	653
R^2	0.000	0.000	0.000	0.000	0.000	0.001	0.001

"Equal rates" conditions ("equal rates" treatment vs. "equal rates" control)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.022	−1.782	0.715	−0.048	0.022	0.049	−0.063*
	(0.039)	(1.387)	(4.426)	(0.038)	(0.032)	(0.038)	(0.033)
Constant	0.509***	49.357***	74.720***	0.429***	0.208***	0.366***	0.264***
	(0.028)	(1.026)	(3.109)	(0.028)	(0.023)	(0.027)	(0.025)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.306$							
Observations	661	661	661	661	661	661	661
R^2	0.000	0.003	0.000	0.002	0.001	0.003	0.006

Disappointment study (treatment vs. control)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	−0.021	0.610	7.267*	0.033	0.008	0.011	−0.064**
	(0.039)	(1.297)	(4.005)	(0.041)	(0.035)	(0.039)	(0.029)
Constant	0.636***	55.844***	62.980***	0.464***	0.245***	0.341***	0.185***
	(0.028)	(0.916)	(2.716)	(0.029)	(0.025)	(0.027)	(0.022)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.214$							
Observations	606	606	606	606	606	606	606
R^2	0.000	0.000	0.005	0.001	0.000	0.000	0.008

Notes: OLS regressions, robust standard errors in parentheses. Within each panel, each column regresses a demographic variable on a treatment dummy to test for imbalanced treatment assignment. In each panel, a joint F-test, estimated in a SUR model, tests the hypothesis that all treatment differences are zero. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.4 Test for balanced treatment assignment – part 2

Leisure time study (treatment vs. control)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	−0.057*	1.020	1.817	−0.010	−0.004	−0.061**	0.033
	(0.030)	(0.805)	(3.236)	(0.030)	(0.024)	(0.030)	(0.024)
Constant	0.515***	37.265***	73.602***	0.563***	0.202***	0.456***	0.172***
	(0.022)	(0.579)	(2.309)	(0.021)	(0.017)	(0.021)	(0.016)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.123$							
Observations	1,095	1,095	1,095	1,095	1,095	1,095	1,095
R^2	0.003	0.001	0.000	0.000	0.000	0.004	0.002

Attention condition (compared to control condition of main study)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Attention	0.011	−1.356	−0.225	−0.042	−0.034	0.064	0.023
	(0.041)	(1.383)	(4.655)	(0.040)	(0.032)	(0.040)	(0.034)
Constant	0.511***	47.116***	76.831***	0.426***	0.213***	0.374***	0.204***
	(0.028)	(0.935)	(3.145)	(0.027)	(0.023)	(0.027)	(0.022)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.400$							
Observations	603	603	603	603	603	603	603
R^2	0.000	0.002	0.000	0.002	0.002	0.004	0.001

Attention "equal rates" condition (compared to "equal rates" control condition)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Attention	−0.026	−2.743*	−3.466	−0.069*	0.002	0.012	−0.009
	(0.041)	(1.472)	(4.485)	(0.040)	(0.034)	(0.040)	(0.036)
Constant	0.509***	49.357***	74.720***	0.429***	0.208***	0.366***	0.264***
	(0.028)	(1.026)	(3.109)	(0.028)	(0.023)	(0.027)	(0.025)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.400$							
Observations	589	589	589	589	589	589	589
R^2	0.001	0.006	0.001	0.005	0.000	0.000	0.000

Notes: OLS regressions, robust standard errors in parentheses. Within each panel, each column regresses a demographic variable on a treatment dummy to test for imbalanced treatment assignment. In each panel, a joint F-test, estimated in a SUR model, tests the hypothesis that all treatment differences are zero. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A.5 Test for balanced treatment assignment – part 3

Counterfactual study (low/high counterfactual condition vs. control)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low count.	−0.018	−1.322	3.011	0.017	0.019	−0.019	0.018
	(0.040)	(1.686)	(4.615)	(0.040)	(0.033)	(0.039)	(0.034)
High count.	−0.046	2.631	0.041	0.059	−0.013	−0.014	0.009
	(0.040)	(2.682)	(4.635)	(0.040)	(0.032)	(0.039)	(0.034)
Constant	0.556***	50.869***	79.513***	0.534***	0.211***	0.393***	0.227***
	(0.028)	(1.389)	(3.218)	(0.028)	(0.023)	(0.028)	(0.024)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.717$							
Observations	945	945	945	945	945	945	945
R^2	0.001	0.003	0.001	0.002	0.001	0.000	0.000

Advantaged counterfactual study (low/high counterfactual condition vs. control)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low count.	0.057	0.658	−5.069	−0.084**	0.046	−0.007	−0.001
	(0.041)	(1.393)	(4.608)	(0.041)	(0.032)	(0.040)	(0.034)
High count.	0.083**	0.967	−0.363	−0.086**	0.012	−0.016	0.049
	(0.041)	(1.395)	(4.648)	(0.041)	(0.031)	(0.040)	(0.035)
Constant	0.481***	48.884***	83.168***	0.557***	0.173***	0.403***	0.226***
	(0.028)	(0.975)	(3.257)	(0.028)	(0.021)	(0.028)	(0.024)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.256$							
Observations	893	893	893	893	893	893	893
R^2	0.005	0.001	0.002	0.007	0.002	0.000	0.003

Vignette study (low/high counterfactual condition vs. control)

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Mid-west	Region: South	Region: West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Low count.	0.009	−0.080	3.792	0.054	−0.039	0.084**	0.011
	(0.034)	(1.209)	(3.637)	(0.035)	(0.030)	(0.034)	(0.026)
High count.	−0.016	−0.976	5.773	0.026	−0.020	0.018	−0.031
	(0.034)	(1.161)	(3.590)	(0.035)	(0.030)	(0.033)	(0.025)
Constant	0.612***	53.918***	66.715***	0.448***	0.249***	0.331***	0.163***
	(0.024)	(0.849)	(2.522)	(0.024)	(0.021)	(0.023)	(0.018)
Joint F-test (H_0: all differences between conditions are zero). $p = 0.083$							
Observations	1,222	1,222	1,222	1,222	1,222	1,222	1,222
R^2	0.000	0.001	0.002	0.002	0.001	0.006	0.002

Notes: OLS regressions, robust standard errors in parentheses. Each column regresses a demographic variable on treatment dummies to test for imbalanced treatment assignment. A joint F-test (SUR model) tests the hypothesis that all treatment differences are zero. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

B Main experiment and robustness experiments

B.1 Treatment effect in main study

Table B.1 Mean treatment effects in main study

Main study: Treatment − Control								
Effort scenario e	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	-1.93	-0.33	-1.58	-1.42	0.29	0.20	1.33	-0.49
Standard error	1.46	1.19	1.28	1.40	1.49	1.32	1.39	0.67
CI, 95%	[-4.8, 0.9]	[-2.7, 2]	[-4.1, 0.9]	[-4.2, 1.3]	[-2.6, 3.2]	[-2.4, 2.8]	[-1.4, 4.1]	[-1.8, 0.8]
p-values, t-tests	0.184	0.781	0.218	0.310	0.848	0.879	0.339	0.464
p-value, F-test	0.668							

Notes: Results from OLS regressions. Columns "0%" to "100%" present results for each of the seven effort scenarios, and Column "Average" presents results averaged across all scenarios. The outcome variable is the reward share assigned to the disadvantaged worker B. "Reward diff." denotes the estimated treatment effect (share in treatment condition versus share in control condition). Robust standard errors, 95% confidence intervals, and p-values are reported. The last row, "p-value, F-test", presents the p-value from an F-test that tests the joint null hypothesis that the differences are zero in each effort scenario. It is estimated in a SUR model with standard errors that are clustered at the respondent level.

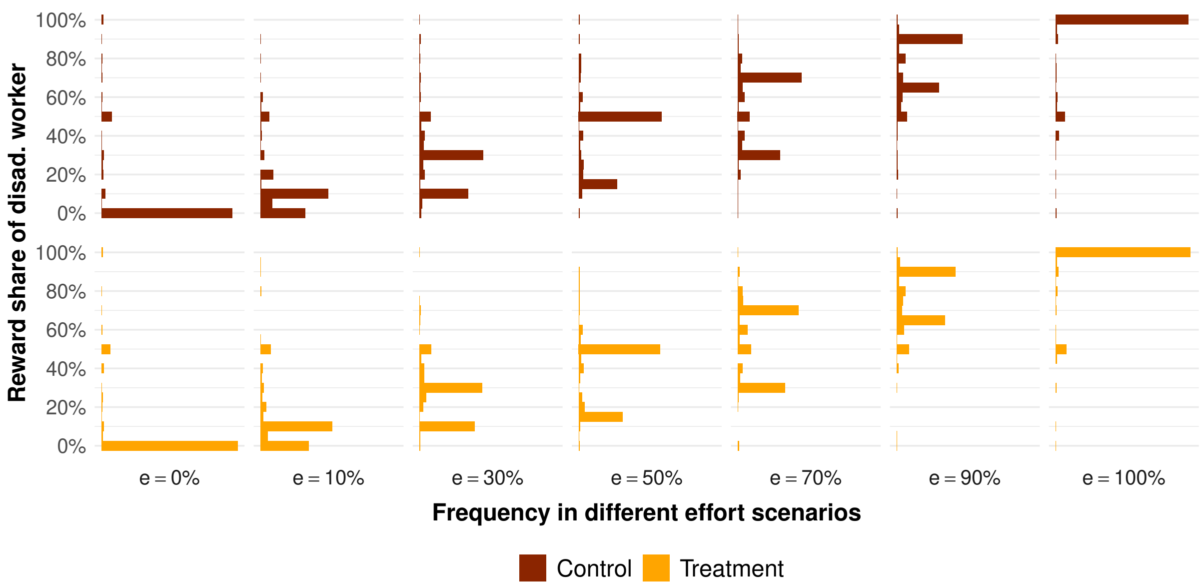


Figure B.1 Histogram of reward share of disadvantaged worker in main study

Notes: Histogram of the reward share assigned to the disadvantaged worker B in the treatment and control condition of the main study.

Table B.2 Heterogeneous treatment effects in the main study

	Mean reward share of disadv. worker (in %)
Treatment	9.953
	(8.966)
Female (bin.)	0.024
	(0.993)
College (bin.)	0.570
	(1.092)
Republican (bin.)	−0.852
	(1.002)
Income (log)	0.180
	(0.621)
Empathy (std.)	0.668
	(0.513)
Internal LOC (std.)	0.467
	(0.458)
Treatment × Female (bin.)	0.448
	(1.389)
Treatment × College (bin.)	−0.336
	(1.495)
Treatment × Republican (bin.)	0.764
	(1.394)
Treatment × Income (log)	−0.993
	(0.832)
Treatment × Empathy (std.)	−0.496
	(0.719)
Treatment × Internal LOC (std.)	−1.571
	(0.656)
Constant	42.098
	(6.663)
Observations	634
R^2	0.019

Notes: Results from the main study, OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share assigned to the disadvantaged worker B, averaged across the seven effort scenarios. The independent variables include interaction terms of the treatment dummy with six respondent characteristics: a dummy for female gender, having a Bachelor's degree, and being Republican, logarithmic income, a standardized empathy score, and a standardized internal locus of control score. p-values of the interaction effects (printed in bold) are adjusted for multiple hypotheses testing with the help of the Benjamini-Hochberg procedure. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.2 Discussion of the contingent response method

Methodological work has explored whether decisions elicited through a contingent response method or strategy method differ systematically from choices elicited through a direct response method. In their review, Brandts and Charness (2011) conclude that most studies do not document such a difference. Moreover, none of the studies reviewed failed to replicate a treatment effect found using a contingent response method with the direct response method. Here, a series of additional analyzes is provided that make clear that this conclusion is very likely to hold in the context of this study, too.

First, the contingent response method requires spectators to believe that each scenario is potentially true. Reassuringly, only a few spectators can distinguish the hypothetical scenarios from the real one, even after seeing all scenarios and making all their redistribution decisions. When asked to guess which of the scenarios is real, 46% respond that they do not know. Among the others, only 16% guess correctly. Therefore, the recognition rate is only slightly higher than what would be expected under random guessing (12.5%).

Second, the experimental results are robust to excluding respondents who recognize the real scenario (see Table B.3 below).

Third, the "leisure time" study, which is introduced in Section 4.3 of the main text and further described in Appendix B.6, replicates the main finding without using a contingent response method. Here, each spectator makes exactly one decision for the real effort choices of one pair of workers.

Table B.3 Treatment effects are robust if spectators who recognize the true scenario are excluded

	Mean reward share of disadvantaged worker (in %)	
	All participants	No recognition of true scenario
	(1)	(2)
Treatment	−0.493	−0.612
	(0.673)	(0.706)
Constant	44.068***	43.950***
	(0.480)	(0.506)
Observations	653	596
R^2	0.001	0.001

Notes: Results from the main study, OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share (in %) a spectator assigns to the disadvantaged worker B, averaged across all seven effort scenarios. The independent variable is a treatment indicator. Column 1 presents the main specification. Column 2 excludes respondents who are able to distinguish the real effort scenario from the hypothetical ones. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.3 Open-text data

In the following, I report three example responses.

> "The reasoning behind the decisions I made was based on how much each worker performed. I had it so that each worker would get the percentage of money according to the percentage of work he completed."
>
> "I believed that each worker should be paid based on their amount of work. The person who completed 45 tasks should make more money than the person who completed 10 tasks, even if they knew they were making less from the beginning."
>
> "The workers agreed to particular rates and amount of work required to be performed. They knowingly made the decisions regarding how much money they earned. Fair or not what you agree to work for is what they should be paid."

Most responses refer to one (sometimes two, rarely more) distinct fairness view. I develop a coding scheme that captures these views. Each response is assigned to the fairness views to which it refers. Table B.4 provides an overview of all codes.

Table B.4 Classification of open-ended responses

Code	Explanation	Example
Fairness codes		
Effort	Reward based on work, effort, task completion.	"People should get paid based on their work quality and effort [...]"
Initial outcome fair	Reward based on initial payments. Workers accepted their work conditions, hence no need for redistribution.	"I based my decisions on the 'ground rules' the workers signed up for before they did the tasks. They each knew that the chance of being paid either \$0.10 or \$0.50 per piece we 50% up front and agreed to do the work."
Equality	Equal rewards irrespective of effort and circumstances.	"I felt that they were equally deserving."
Endogeneity	Acknowledgment that workers' effort choices are shaped by their circumstances.	"[...] The lower rate does not promote someone to work that hard."
Need	Decision shaped by concern that workers need a sufficient income.	"They both need money to survive in this planet"
Residual codes		
Misunderstanding	Explanation clearly reveals misunderstanding of the instructions.	"I'm not quite sure I understood if I was supposed to change the amount paid."
Other	Explanation too vague or nonsensical to assign a fairness code.	"Based on my idealogy of fairness, worker's wages and/or ability."

Notes: This table provides an overview of the different categories in the coding scheme, an explanation for each code, and example extracts from open-text responses that belong to the corresponding category.

Table B.5 Frequency of fairness motives in open-text data

Code	Control	Treatment	p-value
Fairness codes			
Effort	58.3%	59.2%	0.915
Initial outcome fair	8.2%	12.9%	0.375
Equality	10.0%	10.3%	0.915
Endogeneity	0.6%	1.0%	0.915
Need	0.3%	0.6%	0.915
Residual codes			
Misunderstanding	1.6%	1.3%	0.915
Other	27.0%	22.2%	0.575
Sample size	319	311	

Notes: This table shows which share of treatment respondents and control respondents mention different fairness motives in their open-text response. See Table B.4 for details on the coding scheme. The last column contains p-values from χ^2-tests which test for the equality of proportions in each row and are adjusted for multiple hypotheses testing, using the Benjamini-Hochberg procedure.

B.4 "Equal rates" conditions

More details on the experimental conditions

"Equal rates" control condition ($0.50 version) Both workers do not know their realized piece rate while making their effort choices. They are aware that their piece rates might either be $0.50 or $0.10 with equal chance. They learn about their realized piece rate ($0.50 for worker A and $0.50 for worker B) only after completing their work.

"Equal rates" treatment condition Both workers do not know their realized piece rate while making their effort choices. Worker A is aware that his piece rate might either be $0.90 or $0.50 with equal chance. Worker B is aware that his piece rate might either be $0.50 or $0.10 with equal chance. They learn about their realized piece rate ($0.50 for worker A and $0.50 for worker B) only after they finish their work.

I ran the "equal rates" conditions together with the main study in June 2020. The study protocol is identical. As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix A). The instructions are available online (https://osf.io/xj7vc/).

Qualifying note: Workers who receive a $0.90 piece rate receive their payments without a redistribution stage. Workers with a $0.10 piece rate are used in a second variant of the "equal rates" control condition in which both workers earn $0.10. To maximize statistical power, I present results in which I pool the $0.50 and the $0.10 control conditions, but the results are virtually identical if I only use the $0.50 control condition described above.

Table B.6 Mean treatment effects in "equal rates" conditions

"Equal rates": Treatment − Control								
Effort scenario e	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	1.51	0.55	0.64	-0.14	-0.69	-1.70	-0.44	-0.04
Standard error	1.43	1.09	0.67	0.18	0.63	1.15	1.19	0.24
CI, 95%	[-1.3, 4.3]	[-1.6, 2.7]	[-0.7, 1.9]	[-0.5, 0.2]	[-1.9, 0.6]	[-4, 0.6]	[-2.8, 1.9]	[-0.5, 0.4]
p-values, t-tests	0.292	0.613	0.336	0.423	0.277	0.140	0.711	0.872
p-value, F-test	0.747							

Notes: Results from OLS regressions. Columns "0%" to "100%" present results for each of the seven effort scenarios, and Column "Average" presents results averaged across all scenarios. The outcome variable is the reward share assigned to the disadvantaged worker B. "Reward diff." denotes the estimated treatment effect (share in treatment condition versus share in control condition). Robust standard errors, 95% confidence intervals, and p-values are reported. The last row, "p-value, F-test", presents the p-value from an F-test that tests the joint null hypothesis that the differences are zero in each effort scenario. It is estimated in a SUR model with standard errors that are clustered at the respondent level.

B.5 Disappointment study

More details on the experimental conditions

Disappointment control condition Both workers have to complete 10 tasks. They do not know their realized piece rate while making their effort choices. They are aware that their piece rates might either be \$0.50 or \$0.10 with equal chance. They learn about their realized piece rate (\$0.50 for worker A and \$0.10 for worker B) only after they finish their work.

Disappointment treatment condition Both workers have to complete 10 tasks. They are informed about their realized piece rate already before they decide how much effort they exert. Therefore, worker A knows about his high rate of \$0.50 and worker B about his low rate of \$0.10 when they decide how many tasks they complete.

I ran the "disappointment" experiment in February 2021 with a convenience sample of US adults recruited with the help of the survey company Lucid. Treatment assignment is balanced across covariates (see Appendix A). The results are robust to the use of post-stratification weights (see Table B.7). The decisions of spectators are probabilistically incentivized. 25 pairs of real workers were randomly assigned to the spectators. The decisions of the selected spectators were implemented. The instructions are available online (https://osf.io/xj7vc/).

Table B.7 Treatment effects in the disappointment study

	Reward share of disadvantaged worker (in %)	
	(1)	(2)
Treatment	−2.202	−0.763
	(1.422)	(2.122)
Constant	36.695***	35.863***
	(0.973)	(1.387)
Weights	–	✓
Observations	606	606
R^2	0.004	0.000

Notes: Results from the disappointment study, OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share assigned to worker B (low piece rate). The independent variable is a treatment indicator. Column 1 reports the unweighted main specification. Column 2 applies post-stratification weights. The weights render the sample representative for the US general population in terms of gender, age, income, education, and census region. I use a raking algorithm (R package `anesrake`) and follow the guidelines of the American National Election Study to calculate the survey weights (Pasek et al., 2014). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.6 Leisure time study

More details on the experiment I ran the "leisure time" experiment in June 2022 with a convenience sample of US adults via Prolific. Treatment assignment is balanced across covariates (see Appendix A). The instructions are available online (https://osf.io/xj7vc/). Workers could choose between two options. They could work on the task and collect email address data for 30 minutes. (I monitor the output to ensure that the worker worked hard.) Or they could enjoy leisure time on YouTube for 30 minutes and watch whichever video seemed most fun to them. (I ask workers to describe which videos they watched to ensure that they did not spend their time differently.) Spectators were randomly assigned to a control or treatment condition.

Control condition: Workers learn about their realized reward for working (£5 or £1, each with 50% chance) only after they make their work/leisure choice and completed their work/leisure time. Their choice is made in the same environment and, hence, comparable.

Treatment condition: Workers learn about their realized reward for working already before they make their work/leisure choice. One worker is encouraged to work knowing that he can earn the high reward; the other worker is discouraged by knowing that he can only earn the low reward.

Since workers choose between two options, there are only four possible effort scenarios: (i) both workers work, (ii) only worker A works, (iii) only worker B works (rare), or (iv) neither of the workers works (no redistribution possible). I focus on the first two scenarios and run the study *without* using a contingent response method. Spectators make one decision for the real effort choices of one pair of workers. The decision is probabilistically incentivized. 50 pairs of real workers were randomly assigned to the spectators, and the decisions of the selected spectators were implemented.

Table B.8 shows the average treatment effect on the reward share of the disadvantaged worker, discussed in the main text. The results are robust to the use of post-stratification weights (Column 2). If I estimate separate results for the two effort scenarios (not reported), I obtain an effect estimate of 2.6 pp (p = 0.061, n = 739) for the scenario in which only the advantaged worker works and an estimate of 0.1 pp (p = 0.944, n = 356) for the scenario in which both workers work. I correct for multiple hypotheses testing using the Benjamini-Hochberg procedure.

Table B.8 Treatment effects in the leisure time study

	Reward share of disadvantaged worker (in %)	
	(1)	(2)
Treatment	1.993	1.637
	(1.407)	(1.938)
Constant	19.741***	18.286***
	(0.996)	(1.364)
Weights	–	✓
Observations	1,095	1,095
R^2	0.002	0.001

Notes: Results from the leisure time study, OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share assigned to worker B. The independent variable is a treatment indicator. Column 1 reports the unweighted main specification. Column 2 applies post-stratification weights. The weights render the sample representative for the US general population in terms of gender, age, income, education, and census region. I use a raking algorithm (R package `anesrake`) and follow the guidelines of the American National Election Study to calculate the survey weights (Pasek et al., 2014). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.7 Pooled test

Table B.9 Pooled test of treatment effects

	Mean reward share of disadvantaged worker (in %)	
	(1)	(2)
Treatment	−0.025	−0.172
	(0.351)	(0.313)
Study FE	✓	✓
Attention included?	–	✓
Observations	10,293	14,080
R^2	0.064	0.050

Notes: Column 1: Pooled results from the main study, the equal rates study, and the leisure time study. Column 2 additionally includes the attention condition and the attention equal rates condition. OLS regressions, standard errors are clustered on the level of participants and reported in parentheses. The outcome variable is the reward share assigned to worker B (low piece rate). The independent variable is a binary treatment indicator. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C Mechanism evidence

C.1 Beliefs about circumstances' effect on choices in the main study

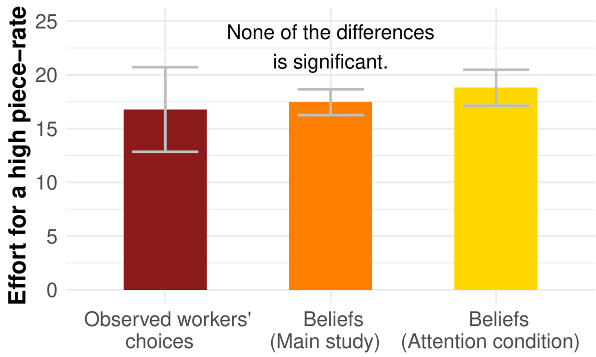


Figure C.1 Average beliefs about the piece-rate effect (with 95% CI)

Notes: Results from the main study and the attention condition. The figure presents the average observed and average perceived effort choices of workers for a high piece rate of $0.50. The average number of completed tasks for a low piece rate is 5.04. Red bar: Actual effort decisions of workers. Orange bar: Effort choice that spectators expect in the main study. Yellow bar: Effort choice that spectators expect in attention condition. The gray errorbars are 95% confidence intervals. t-tests are used to evaluate the significance of the differences.

Table C.1 Treatment effects among respondents who believe in piece-rate effect

	Mean reward share of disadvantaged worker (in %)				
	(1)	(2)	(3)	(4)	(5)
Treatment	−0.493	−0.368	−0.077	−0.278	−0.030
	(0.673)	(0.833)	(0.856)	(1.071)	(1.701)
Constant	44.068***	44.064***	43.925***	44.846***	44.479***
	(0.480)	(0.583)	(0.609)	(0.773)	(1.254)
Perceived incentive effect	–	>1	≥2	≥4	≥6
Observations	653	396	373	222	98
R^2	0.001	0.000	0.000	0.000	0.000

Notes: Results from the main study, OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share (in %) a spectator assigns to the disadvantaged worker B, averaged across all seven effort scenarios. The independent variable is a treatment indicator. Column 1 presents the main specification. Columns 2-5 report regressions for subsamples with increasingly higher perceived incentive effects (as indicated in the row "Perceived incentive effect"). For example, Column 3 restricts the sample to respondents who believe that workers complete at least twice as many tasks for the high than for the low piece rate. Belief data are available for 540 respondents. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C.2 Attention manipulation

"Equal rates" attention condition Panel (C) of Table C.2 builds on an analogous attention manipulation that extends the "equal rates" conditions. I explicitly inform spectators that "the piece rates strongly influence the number of tasks a worker completes." Spectators learn how large this incentive effect is on average (in the equal rates conditions, see Appendix E) and read two typical comments by workers that explain why this is the case. Participants must spend at least 20 seconds on this information page, whose key message is repeated on the next page and tested for in the subsequent quiz. Data for this condition were collected together with the main study, the "equal rates" conditions, and the main attention manipulation discussed in the main text.

Table C.2 Mean treatment effects of attention manipulation

(A) Attention manipulation: Attention − Control (compared to main control condition)

Effort scenario e	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	-1.24	0.88	-0.88	-1.28	-1.38	-0.14	0.04	-0.57
Standard error	1.52	1.31	1.40	1.48	1.52	1.40	1.53	0.72
CI, 95%	[-4.2, 1.7]	[-1.7, 3.4]	[-3.6, 1.9]	[-4.2, 1.6]	[-4.4, 1.6]	[-2.9, 2.6]	[-3, 3]	[-2, 0.8]
p-values, t-tests	0.412	0.504	0.529	0.388	0.366	0.921	0.980	0.423
p-value, F-test	0.583							

(B) Attention manipulation: Attention − Treatment (compared to main treatment condition)

Effort scenario e	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	0.69	1.21	0.70	0.14	-1.66	-0.34	-1.29	-0.08
Standard error	1.41	1.32	1.33	1.46	1.52	1.33	1.45	0.71
CI, 95%	[-2.1, 3.5]	[-1.4, 3.8]	[-1.9, 3.3]	[-2.7, 3]	[-4.6, 1.3]	[-2.9, 2.3]	[-4.1, 1.6]	[-1.5, 1.3]
p-values, t-tests	0.626	0.360	0.601	0.923	0.275	0.799	0.374	0.910
p-value, F-test	0.768							

(C) "Equal rates" attention man.: Attention − Control (compared to "equal rates" control condition)

Effort scenario e	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	-1.73	-0.03	0.03	0.16	-0.38	-0.47	-0.60	-0.43
Standard error	1.29	1.14	0.75	0.21	0.73	1.20	1.27	0.23
CI, 95%	[-4.3, 0.8]	[-2.3, 2.2]	[-1.4, 1.5]	[-0.2, 0.6]	[-1.8, 1]	[-2.8, 1.9]	[-3.1, 1.9]	[-0.9, 0]
p-values, t-tests	0.178	0.979	0.968	0.452	0.601	0.698	0.638	0.066
p-value, F-test	0.208							

Notes: Results from OLS regressions. Each panel presents the results from a different comparison of experimental conditions. The title of each panel describes which experimental conditions are compared. Columns "0%" to "100%" present results for each of the seven effort scenarios, and Column "Average" presents results averaged across all scenarios. The outcome variable is the reward share assigned to the disadvantaged worker B. "Reward diff." denotes the estimated treatment effect. Robust standard errors, 95% confidence intervals, and p-values are reported. The last row, "p-value, F-test", presents the p-value from an F-test that tests the joint null hypothesis that the differences are zero in each effort scenario. It is estimated in a SUR model with standard errors that are clustered at the respondent level.

C.3 Counterfactual study

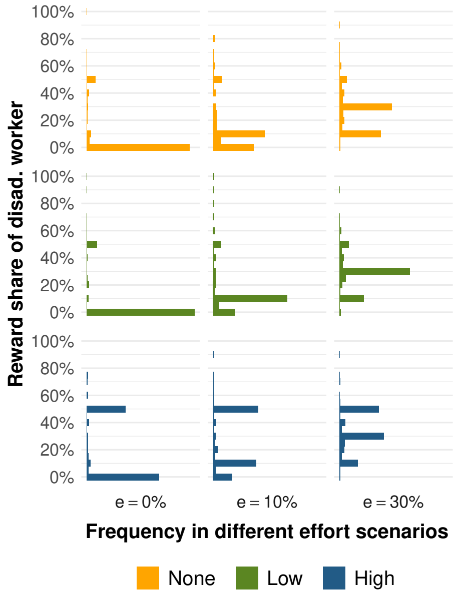


Figure C.2 Counterfactual study: Histograms of reward share of disadv. worker

Notes: Histograms of the reward share assigned to the disadvantaged worker B for each experimental condition and each effort scenario in the counterfactual study.

Table C.3 Mean treatment effects in counterfactual study

(A) Low counterfactual − No information

Effort scenario e	0%	10%	30%	Average
Reward diff.	-0.13	1.58	3.32	1.59
Standard error	1.34	1.31	1.11	1.03
CI, 95%	[-2.8, 2.5]	[-1, 4.1]	[1.1, 5.5]	[-0.4, 3.6]
p-values, t-tests	0.923	0.227	0.003	0.123
p-value, F-test	0.011			

(B) High counterfactual − No information

Effort scenario e	0%	10%	30%	Average
Reward diff.	12.31	12.75	8.69	11.25
Standard error	1.65	1.49	1.21	1.23
CI, 95%	[9.1, 15.5]	[9.8, 15.7]	[6.3, 11.1]	[8.8, 13.7]
p-values, t-tests	<0.001	<0.001	<0.001	<0.001
p-value, F-test	<0.001			

Notes: Counterfactual study, results from OLS regressions. Panel A compares the *Low counterfactual* with the *No information* condition. Panel B compares the *High counterfactual* with the *No information* condition. Columns "0%" to "30%" present results for each of the three effort scenarios, and Column "Average" presents results averaged across all three scenarios. The outcome variable is the reward share assigned to the disadvantaged worker B. "Reward diff." denotes the estimated treatment effect. Robust standard errors, 95% confidence intervals, and p-values are reported. The last row, "p-value, F-test", presents the p-value from an F-test that test the joint null hypothesis that the differences are zero in each effort scenario. It is estimated in a SUR model with standard errors that are clustered at the respondent level.

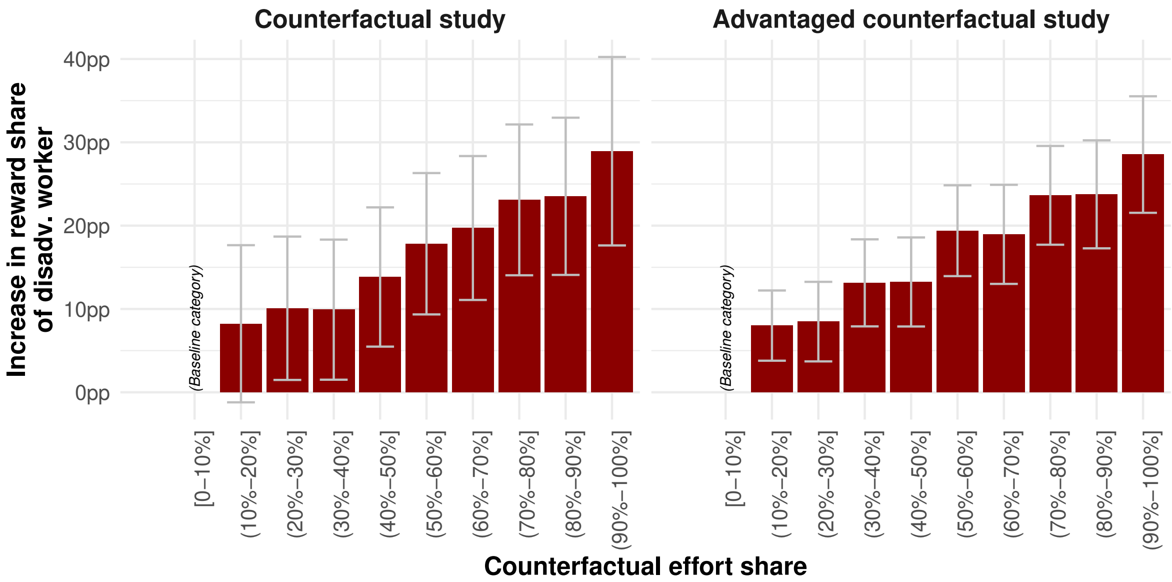


Figure C.3 Counterfactual study, scenarios 4–7: Reward decisions react to counterfactual effort share

Interpretation: The reward share assigned to the disadvantaged worker increases with his counterfactual effort share: the higher the counterfactual effort share, the higher the reward share.
Notes: Results from the counterfactual study (left panel) and the advantaged counterfactual study (right panel). Decision data from scenarios 4–7 with randomly generated effort choices and counterfactual effort choices. Coefficient estimates from OLS regressions with 95% confidence intervals (standard errors clustered on participant level) are displayed. I regress the reward share given to the disadvantaged worker B on dummies indicating whether worker B's counterfactual effort share falls in the interval (10%-20%], (20%-30%], etc., while flexibly controlling for the share of actual effort.

C.4 Advantaged counterfactual study

I ran the "advantaged counterfactual" experiment in June 2022 with a US sample that I recruited with the help of the survey company Lucid. Treatment assignment is balanced across covariates (see Appendix A). The instructions are available online (https://osf.io/xj7vc/). Spectators are randomly assigned to one of three experimental conditions.

No information condition (short: None): No information is provided about worker A's counterfactual effort choice.

Low counterfactual condition (short: Low): Spectators learn that worker A would complete as few tasks as worker B for a low piece rate. Hence, workers A and B (would) make the same choices in the disadvantaged environment.

High counterfactual condition (short: High): Spectators learn that worker A would not change his effort provision and thus would not exert less effort for a lower piece rate. This basically means that worker A's effort choice is not shaped by his circumstances.

The seven hypothetical effort scenarios are designed analogously to the main counterfactual study (Table 5).

The effect from the main counterfactual study (Figure 3) should reverse. A "low counterfactual" choice of an advantaged worker implies that both workers behave identically under equal circumstances and therefore corresponds to the "high counterfactual" treatment in the main counterfactual study. Likewise, a "high counterfactual" choice of an advantaged worker implies that the workers still behave differently under equal circumstances and hence corresponds to the "low counterfactual" treatment in the main counterfactual study. Figure C.4 shows that the effects reverse and replicate the findings of the main counterfactual study.

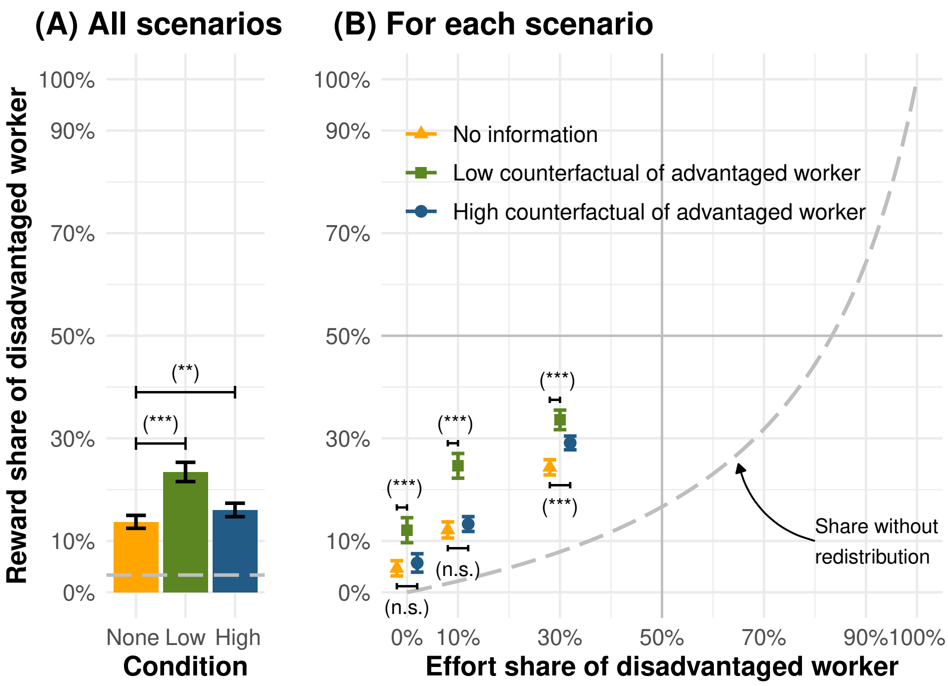


Figure C.4 Advantaged counterfactual study: Mean reward share of disadv. worker (with 95% CI)

Notes: Results from the "advantaged counterfactual" study, decisions 1-3. Panel A displays the mean reward share assigned to the disadvantaged worker B in each experimental condition, averaged across all three effort scenarios, with 95% confidence intervals. Panel B plots the mean reward share in each effort scenario with 95% confidence intervals. The gray dashed line shows the default share, that is, which payment share worker B would receive if spectators do not redistribute. I test for differences between the "High counterfactual" and the "No information" condition (upper test) and between the "Low counterfactual" and the "No information" condition (lower test). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, (n.s.) $p \geq 0.10$.

C.5 Rationale study

More details on the study I conducted the rationale study in September 2022 with a convenience sample of US adults recruited with the help of the survey platform Prolific. Respondents receive a hypothetical version of the counterfactual study with only one effort scenario, namely scenario 3 of the counterfactual study. After indicating how they would distribute the rewards in this situation, they respond to the following open-ended survey question: "Please explain why you made your decision." Screenshots of the instructions are available online (https://osf.io/xj7vc/).

Coding scheme Responses are manually classified into the following categories.

Actual choice meritocratism: Participants explain to reward workers (mainly) for the amount of tasks actually completed.

Comparable choice meritocratism: Participants explain that they also consider worker B's counterfactual effort choice and/or indicate that they compensate him for the bad luck of a low piece rate that discourages effort.

Other fairness argument: The response cannot be clearly assigned to one of the categories above. For example, the response could be too vague to clearly distinguish between actual choice and comparable choice meritocratism or it could refer to another fairness ideal.

C.6 Structural model of fairness views

Data Counterfactual study, decisions 4–7, 630 respondents, conditions: high counterfactual and low counterfactual. In decisions 4–7, respondents face a randomly generated effort scenario.[28] The effort share of worker B and his counterfactual effort share (had he earned a high piece rate) are drawn as follows.

- Effort of worker A: Uniformly randomly drawn from the set $\{0, 1, ..., 50\}$.
- Effort of worker B (E_B): Uniformly randomly drawn from the set $\{0, 1, ..., 25\}$.
- Counterfactual effort of worker B for a high piece rate: Uniformly randomly drawn from the set $\{E_B, E_B + 1, ..., 50\}$.
- The effort and initial payment shares of both workers follow from the above variables.

Model The model has five parameters: the population shares $\boldsymbol{\theta}$ of the four merit views ($\sum_t \theta_t = 1$) and the standard deviation of the response error σ. I impose $0 \leq \theta_t \leq 1 \; \forall t$ and $\sigma > 0$. The log-likelihood is described by:

Log-likelihood

[28]The contingent response method allows me to freely vary the effort choices of workers in the hypothetical scenarios without deceiving participants.

$$\log F(\boldsymbol{r} \mid \boldsymbol{\theta}, \sigma) = \sum_i \log f_i(\boldsymbol{r}_i \mid \boldsymbol{\theta}, \sigma) \tag{1}$$

$$f_i(\boldsymbol{r}_i|\boldsymbol{\theta}, \sigma) = \sum_t \theta_t f_i^t(\boldsymbol{r}_i \mid \theta_t, \sigma) \tag{2}$$

$$f_i^t(\boldsymbol{r}_i \mid \theta_t, \sigma) = \prod_s \varphi(r_{is} - m_i^t(s), \sigma^2) \tag{3}$$

where φ denotes the normal density function.

Estimation I estimate the model in R with the help of the `maxLik` package (Henningsen and Toomet, 2011). The BFGS algorithm is used to solve the constrained optimization problem. I estimate σ and the share of actual choice meritocrats, comparable choice meritocrats, and libertarians. The share of egalitarians follows through $\sum_t \theta_t = 1$.

Computational robustness I confirm the numerical stability of the maximum likelihood estimator in three steps. First, I replicate the results in 100 estimations with random start parameters. Second, I generate 100 simulated data sets from the model with randomly drawn parameters and confirm that the estimates recover the parameters of the models. Third, I replicate the results with the Nelder-Mead optimization algorithm.

Robustness of estimates Table C.4 shows that the results of the maximum likelihood are robust in several different specifications.

- **Duration**: Excludes respondents with a response duration that is lower than the 25% percentile.
- **Quiz**: Excludes respondents who answer at least one quiz question wrongly.
- **Guess correct**: Excludes respondents who are able to distinguish the real scenario from the hypothetical ones.
- **Scenarios 1–7**: Also includes decision data from scenarios 1–3.
- **Bounds adjust**: Because the support of normal noise is unbounded, the likelihood function assigns a positive probability to reward shares below 0% or above 100% that cannot occur in practice. Here, I limit the support to values that can occur in practice. I rescale each error density by the inverse cumulative density that lies outside the interval [0%-100%].
- **U-Noise type**: The standard model imposes that all individuals can be classified into four distinct fairness types, from which they randomly deviate by $\varepsilon_{is} \sim_{iid} N(0, \sigma^2)$. The model does not allow for a "pure" noise type that responds in a purely random way. In this robustness check, I introduce such a fifth pure noise type whose reward decisions are i.i.d. uniformly drawn from the full support: $r_{is} \sim_{iid} U[0, 100]$.

However, I cannot credibly identify two separate sources of noise: first, the random deviations of the four fairness types (σ); second, the population share of the pure noise type (θ_{noise}). Therefore, I impose $\sigma = 9.58$ (as estimated in the main specification) and test whether I obtain similar estimates for the population shares if I additionally allow for the pure noise type. The model estimates a population share of 10% for the pure noise type. Reassuringly, the estimated shares of actual choice meritocrats and comparable choice meritocrats are slightly smaller but still similar to the main results.

- **Trembling**: I explore an alternative error specification. The respondents have a "trembling hand" and their response r_{is} is fully random (uniform over [0%-100%]) with probability α. With probability $1 - \alpha$, their response is very close to their merit view (normal error with a standard deviation of 2 percentage points).

Heterogeneity by demographics The model allows to estimate whether its parameters differ for subgroups of respondents. Consider two groups of respondents, group A and group B. I assume that the population shares of different fairness types are $\boldsymbol{\theta}$ in group A. In group B, the population shares are $\boldsymbol{\theta} + \boldsymbol{\lambda}$. That is, I allow each parameter p to differ by λ_p between both groups.

I estimate this model separately for the following group comparisons: male versus female respondents, respondents with below-median versus above-median income, respondents without versus with college degree, Democrats versus Republicans. Table C.5 displays the resulting estimates of $\boldsymbol{\lambda}$.

Table C.4 Robustness of structural estimation

	(1) **Main**	(2) **Duration**	(3) **Quiz**	(4) **Guess correct**	(5) **Scenarios 1-7**	(6) **Bounds adjust**	(7) **U-noise type**	(8) **Trembling**
Population shares								
Actual choice meritocrats	40.0%	38.7%	43.3%	40.0%	44.0%	39.4%	37.6%	37.5%
	(2.1%)	(2.3%)	(2.5%)	(2.2%)	(2.1%)	(2.1%)	(2.1%)	(2.1%)
Comparable choice meritocrats	28.4%	31.4%	29.2%	28.3%	27.9%	28.4%	26.6%	31.8%
	(1.9%)	(2.2%)	(2.3%)	(2.1%)	(1.9%)	(1.9%)	(1.9%)	(2.0%)
Libertarians	16.2%	17.1%	16.1%	16.6%	14.8%	17.0%	14.9%	18.2%
	(1.5%)	(1.7%)	(1.8%)	(1.6%)	(1.5%)	(1.6%)	(1.5%)	(1.6%)
Egalitarians	15.4%	12.8%	11.4%	15.0%	13.3%	15.1%	11.4%	12.5%
	(–)	(–)	(–)	(–)	(–)	(–)	(1.4%)	(–)
Error term and sample								
σ noise	9.58	9.38	9.06	9.63	11.50	9.97	9.58	
	(0.14)	(0.15)	(0.16)	(0.15)	(0.12)	(0.16)	(–)	
α noise								0.23
								(0.01)
Respondents	630	496	434	554	630	630	630	630
Decisions	2520	1984	1736	2216	4410	2520	2520	2520

Notes: Results from counterfactual study, decisions 4–7. Maximum likelihood estimation of the structural model of fairness views. Standard errors in parentheses. The estimates indicate the population shares of different fairness views. The columns estimate the model for different specifications. See text above. No standard errors are reported for the share of egalitarians because their share is deduced from the other estimates (except for Column 7).

Table C.5 Differences in model parameters ($\boldsymbol{\lambda}$) by group

	(1) **Female** (vs. male)	(2) **Income >median** (vs. ≤median)	(3) **College degree** (vs. none)	(4) **Republican** (vs. Democrats)
Differences in shares				
Actual choice meritocrats	7.8%*	4.2%	2.1%	-0.1%
	(4.2%)	(4.2%)	(4.2%)	(4.2%)
Comparable choice meritocrats	-1.1%	5.9%	6.2%	-3.5%
	(3.9%)	(3.9%)	(3.9%)	(3.9%)
Libertarians	-4.2%	-5.6%*	-5.8%*	0.5%
	(3.1%)	(3.0%)	(3.2%)	(3.1%)
Egalitarians	-2.5%	-4.5%	-2.5%	3.1%
	(–)	(–)	(–)	(–)
Sample				
Respondents	611	611	611	611
Decisions	2444	2444	2444	2444

Notes: Results from counterfactual study, decisions 4–7. Maximum likelihood estimation of the structural model of fairness views which allows for different parameters across two groups of individuals. Standard errors in parentheses. The table reports the estimated differences in parameters ($\boldsymbol{\lambda}$). For the sake of brevity, the baseline estimates ($\boldsymbol{\theta}$) as well as the normal error (σ, constant across groups) are not reported. The columns report results from separate estimations. The column labels indicate which two demographic groups are compared. See text above. No standard errors are reported for the share difference of egalitarians because their share is deduced from the other estimates. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

D Vignette study

Table D.1 Robustness of the results from the vignette study

(A) Share of respondents redistributing towards the disadvantaged worker

	Binary indicator for compensation			
	Main	Keep 45s+	Keep 75s+	Weighted
	(1)	(2)	(3)	(4)
Low counterfactual	−0.004	−0.016	0.002	−0.000
	(0.029)	(0.029)	(0.031)	(0.038)
High counterfactual	0.126***	0.122***	0.122***	0.135***
	(0.029)	(0.029)	(0.031)	(0.037)
Vignette FE	✓	✓	✓	✓
Observations	2,664	2,789	2,390	2,664
R^2	0.028	0.028	0.027	0.024

(B) Mean reward share of disadvantaged person

	Reward share of disadv. person (in %)			
	Main	Keep 45s+	Keep 75s+	Weighted
	(1)	(2)	(3)	(4)
Low counterfactual	−1.539	−1.828*	−0.974	−1.332
	(1.085)	(1.075)	(1.121)	(1.495)
High counterfactual	6.795***	6.921***	6.861***	6.847***
	(1.177)	(1.175)	(1.224)	(1.447)
Vignette FE	✓	✓	✓	✓
Observations	2,664	2,789	2,390	2,664
R^2	0.135	0.133	0.139	0.116

Notes: Results from the vignette study. OLS regressions with standard errors clustered at the respondent level. The dependent variable in Panel A is a binary indicator for whether a respondent compensates the disadvantaged person by redistributing money towards him. The dependent variable in Panel B is the reward share assigned the disadvantaged person. The independent variables are treatment dummies. Columns 1 shows the main specification. Column 2-4 report different robustness checks. *Keep 45s+*: Exclude respondents who complete the vignettes with an average response time of less than 45 seconds (instead of 60s). *Keep 75s+*: Exclude respondents who complete the vignettes with an average response time of less than 75 seconds (instead of 60s). *Weighted*: Weighted OLS regression with survey weights that render the sample representative for the US general population in terms of gender, age, income, education, and census region. I use a raking algorithm (R package `anesrake`) and follow the guidelines of the American National Election Study to calculate the survey weights (Pasek et al., 2014). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

E How circumstances shape effort in the worker setting

This appendix documents that the piece rates strongly influence how much effort a worker exerts. I study the effort choices of 861 workers who were recruited for the study. 336 workers were recruited for the main study and the additional "equal rates" and attention conditions (Amazon Mechanical Turk, US, May and June 2020). 424 were recruited for the counterfactual study and the advantaged counterfactual study (Amazon Mechanical Turk, US, January 2021 and June 2022). 101 were recruited for the leisure time study (Prolific, US, June 2022).[29]

Table E.1 regresses workers' effort on an indicator for a high piece rate. Specifically,

- Column 1, Main: "High rate" means a piece rate of \$0.50 instead of \$0.10.
- Column 2, Equal rates: "High rate" means (uncertain) piece-rate prospects of \$0.50 or \$0.90 (with equal chance) instead of \$0.10 or \$0.50 (with equal chance).
- Column 3, Counterfactual: "High rate" means a piece rate of \$0.50 instead of \$0.10. The counterfactual study uses a within-subject design. Each worker decides how much effort he would exert for a high piece rate and for a low piece rate.
- Column 4, Leisure time: "High rate" means that workers can earn £5 instead of £1 for 30 minutes of work. A within-subject design is used. Each worker decides whether he would work for the high and for the low rate.

The higher piece rate leads to a 333% higher effort in the main condition, a 155% higher effort in the "equal rates" condition, a 337% higher effort in the counterfactual studies, and a 293% higher effort in the leisure time study. Thus, the external piece rate strongly affects how much effort the workers exert.

Table E.1 The effect of a high piece rate on workers' effort

	Number of completed tasks			**Decides to work**
	Main	Equal rates	Counterfactual	Leisure time
	(1)	(2)	(3)	(4)
High rate	11.744***	5.553**	12.075***	0.335***
	(2.308)	(2.357)	(1.134)	(0.088)
Constant	5.040***	10.044***	5.099***	0.174***
	(1.135)	(1.226)	(0.710)	(0.056)
Observations	124	212	424	101
R^2	0.142	0.029	0.147	0.121

Notes: OLS regressions, robust standard errors in Columns 1 and 2, standard errors clustered at the worker level in Columns 3 and 4. The dependent variable is the number of tasks a worker completes. "High rate" is an indicator for high piece rate (prospects).

[29] In addition, I recruited 56 workers for the *robustness "disappointment"* study (Amazon Mechanical Turk, US, February 2021). Workers in this condition do not make an effort choice. They have to complete exactly ten tasks.

F Research transparency

Preregistration The main study, the “equal rates” conditions, the attention condition, the “equal rates” attention condition, the disappointment study, the “leisure time” study, the counterfactual study, and the “advantaged counterfactual” study were preregistered as project #AEARCTR-0005811 at the AEA RCT Registry. The preregistration includes details on the experimental design, the full experimental instructions, thus the full list of measured variables, the sampling process and planned sample size, exclusion criteria, hypotheses, and the main analyses. The following notes document where I deviate from the preregistration.

- The preregistration occasionally uses different titles, study names, and treatment labels.
- Non-preregistered analyses include the comparison of worker B’s reward share, averaged across effort scenarios (a straightforward summary of the scenario-by-scenario differences), and the structural estimation.
- Wherever I explicitly deviate from the analysis plan, I choose a more conservative approach. For instance, I do not adjust the treatment comparisons in each effort scenario for multiple hypothesis testing. This renders their non-significance even more conservative. The highly significant effects in the counterfactual study survive even conservative adjustments for multiple hypotheses testing.
- The sample sizes differ slightly from the preregistered sample sizes due to the logistics of the sampling process and the preregistered exclusion criteria, which were only applied after the data collection.
- The preregistration defines the difference in payment shares $\Delta p = \frac{P_A}{P_A+P_B} - \frac{P_B}{P_A+P_B}$ as the main outcome variable. In contrast, I use worker B’s payment share $p = \frac{P_B}{P_A+P_B}$ as main outcome variable. Since both are linearly dependent ($p = \frac{1-\Delta p}{2}$), this difference does not affect the results but eases their interpretation.

The rationale study, the vignette study, and the vignette evaluation study were not preregistered.

Ethics approval The study obtained ethics approval from the German Association for Experimental Economic Research (#HyegJqzx, 12/11/2019).

Data and code availability All data and code will be made available online.

Competing interests I declare that I have no competing interests.

G Extract from the instructions of the main study

This appendix shows the central experimental instructions from the main study. The complete experimental instructions for all studies are available at https://osf.io/xj7vc/.

The context of your decision

Our institute currently hires adults from the US general public on an online job portal to work on an important task for one of our projects.

Task
These workers search for publicly available email addresses of academic economists. In each task, a worker is given the name of one economist, searches for the economist's personal or university webpage, identifies his or her email address and sends it to us.

The task requires no special qualification or ability, but demands concentration and effort. Typically, it takes about 2 minutes to complete one task.

Workers can freely choose how long they work and how many tasks they want to complete. At most, they can complete 50 tasks.

– PAGE BREAK –

The context of your decision

Payment
Each worker receives a fixed reward of $1.00 for completing the job as well as a variable payment. The variable payment depends on the number of completed tasks, a piece-rate, and your decisions in this survey. From now on, when we say "payment", we are only referring to this variable payment. It is calculated in two steps:

(1) A worker initially earns a fixed amount for each solved task. We refer to this amount per task as a piece-rate.

variable payment = *number of tasks* x *piece-rate*

For example, a worker who has a piece-rate of $0.20 and solves 10 tasks receives a variable payment of $2 (namely $0.20 x 10).

(2) Afterwards, someone else determines the final payments. Workers are informed about this, although they do not know how and why this happens.

This is where you come into play ...

Your decisions

In the last weeks, we hired 200 workers and matched them into 100 pairs. The decisions that you and others make in this study determine their final earnings. We randomly select one study respondent for each pair of workers.

If you are one of the selected respondents, **your decisions determine the final earnings of a pair of workers**. Let us call them *worker A* and *worker B*.

You can redistribute the payments between worker A and worker B. That is, you decide which share of the total payment amount each worker receives.

Example: Worker A receives a payment of $10 and worker B of $5 so that the sum of their payments is $15. You can freely choose how to distribute the total amount of $15 between both workers.

Completely anonymous: Please note that your decisions are completely anonymous. The workers will receive the shares that you choose with no further information. In particular, they will not learn anything about you or the nature of your decisions.

– PAGE BREAK –

Multiple decisions - each might matter

We ask you to consider **8 different scenarios** corresponding to different possible work outcomes for worker A and worker B. 7 of those scenarios are hypothetical. 1 scenario is real and describes what actually happened when worker A and worker B worked on this task.

You will make **one distribution decision for each scenario**. If you are among the selected respondents, your decision in the real scenario is implemented and determines how much each worker earns. However, you will not be told which scenario really happened, so all of your decisions are important.

Therefore, please take each decision seriously. It might matter a lot to two real workers from the US.

– PAGE BREAK –

The piece-rates

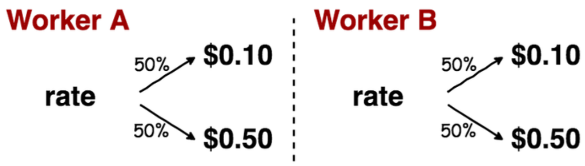
Recall that the piece-rates of the workers determine how much they initially earn for each task. In what follows, we explain how these piece-rates are determined.

The piece-rates

Please read the following information very carefully.

The piece-rate of each worker was determined randomly by a virtual coin flip. Each worker had a 50% chance to get a piece-rate of $0.10 and a 50% chance to get a piece-rate of $0.50. One coin flip determined the rate of worker A, and another coin flip determined the rate of worker B.

Thus, the workers had equal prospects to work for the low or the high rate.



– INFORMATION FOR *CONTROL* GROUP –

Importantly, workers did not know during their work which piece-rate they would get. Only the chances of getting the rates were known. The coin was flipped only after a worker completed and submitted the job. Only then, a worker was informed about his or her definite piece-rate.

In the end, the coin flip determined the following definite rates:

- **Worker A** had a rate of **$0.50**.
- **Worker B** had a rate of **$0.10**.

Thus, they worked for a different rate, but they were informed about their rate only after they completed the job.

– INFORMATION FOR *TREATMENT* GROUP –

Importantly, workers knew which piece-rate they would get before starting their work. The coin was flipped before the workers started working and workers were informed about the result directly.

The coin flip determined the following definite rates:

- **Worker A** had a rate of **$0.50**.
- **Worker B** had a rate of **$0.10**.

Thus, they worked for a different rate.

– EXAMPLE: REDISTRIBUTION DECISION FOR *CONTROL* GROUP –

	Rate prospects (known to worker)	Final rate (unknown to worker)	Completed tasks	Initial payment
Worker A	$0.10 or $0.50 50% chance for each	$0.50	**45 tasks** 90% of total work	**$22.50** 98% of total payment
Worker B	$0.10 or $0.50 50% chance for each	$0.10	**5 tasks** 10% of total work	**$0.50** 2% of total payment
			Total payment:	$23.00

Please split the total payment between both workers.
To do so, please specify which share of the total payment each worker gets. The shares need to add up to 100%.

Share of **worker A**	0	%
Share of **worker B**	0	%
Total	0	%

– EXAMPLE: REDISTRIBUTION DECISION FOR *TREATMENT* GROUP –

	Rate (known to worker)	Completed tasks	Initial payment
Worker A	$0.50	**45 tasks** 90% of total work	**$22.50** 98% of total payment
Worker B	$0.10	**5 tasks** 10% of total work	**$0.50** 2% of total payment
		Total payment:	$23.00

Please split the total payment between both workers.
To do so, please specify which share of the total payment each worker gets. The shares need to add up to 100%.

Share of **worker A**	0	%
Share of **worker B**	0	%
Total	0	%