Discussion Paper Series – CRC TR 224

Discussion Paper No. 318
Project A 01

Shallow Meritocracy
An Experiment on Fairness Views

Peter Andre [1]

September 2021

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

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

Shallow Meritocracy

An Experiment on Fairness Views

Peter Andre
University of Bonn, p.andre@uni-bonn.de
September 2, 2021

Abstract: Meritocracies aspire to reward effort and hard work but promise not to judge individuals by the circumstances they were born into. The choice to work hard is, however, often shaped by circumstances. This study investigates whether people's merit judgments are sensitive to this endogeneity of choice. In a series of incentivized experiments with a large, representative US sample, study participants judge how much money two workers deserve for the effort they exerted. In the treatment condition, unequal circumstances strongly discourage one of the workers from working hard. Nonetheless, I find that individuals hold the disadvantaged worker fully responsible for his choice. They do so, even though they understand that choices are strongly influenced by circumstances. Additional experiments identify the cause of this neglect. In light of an uncertain counterfactual state – what would have happened on a level playing field – participants base their merit judgments on the only reliable evidence they possess: observed effort levels. I confirm these patterns in a structural model of merit views and a vignette study with real-world scenarios.

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

Keywords: Meritocracy, attitudes toward inequality, redistribution, fairness, responsibility, social preferences, inference, uncertain counterfactual.

Acknowledgements: I thank Felix Chopra, Thomas Dohmen, Armin Falk, Thomas Graeber, Leander Heldring, Luca Henkel, Paul Hufe, Ingo Isphording, Fabian Kosse, Yucheng Liang, Matt Lowe, Wladislaw Mill, Franz Ostrizek, Christopher Roth, Sebastian Schaube, Andreas Stegmann, Florian Zimmermann, and participants at various conferences 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 pre-registratered at the AEA RCT Registry (#AEARCTR-0005811). Data and code will be made available. I declare no competing interests. See also Appendix D on research transparency. **Instructions**: The full 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 we consider to be fair, which redistributive policies we implement, and how we design our welfare states (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. Besides talent and skill, the choice to work hard and exert effort is considered a central determinant of merit. By contrast, external circumstances outside the individual's control, such as parental background, race, or sex, are not legitimate sources of merit (Alesina and Angeletos, 2005; Almås et al., 2020; Cappelen et al., 2020b; Konow, 2000). Meritocratic fairness thus distinguishes between effort choices (relevant for merit) and external circumstances (irrelevant for merit). However, this distinction is clouded by a fundamental feature of reality: Agents' choices are endogenous to and shaped by their circumstances, opportunities, and incentives. For instance, a person growing up with few opportunities and incentives to work hard might respond by exerting little effort. Likewise, minorities that experience discrimination might be discouraged from working hard. Indeed, empirical studies have linked effort, career, and schooling choices to gender norms, racial inequality, and the socio-economic environment (e.g., Altmejd et al., 2021; Bursztyn et al., 2017; Carrell et al., 2010; Falk et al., 2020a; Glover et al., 2017). Moreover, the fact that adverse environments often encourage detrimental decision-making is considered a key cause of poverty (e.g., Bertrand et al., 2004; Haushofer and Fehr, 2014).

Any meritocracy thus needs to take a stance on how choices that are shaped by external circumstances should be rewarded. Should choices be evaluated in light of or irrespective of their circumstances? This study explores the prevailing concept of meritocratic fairness and investigates how people reward choices in a series of online experiments with a large, representative US sample of about 4,000 respondents. The study proceeds in four steps. First, I isolate and identify the effect of interest, which requires the control of an incentivized choice experiment. I find that merit judgments completely neglect the endogeneity of choices. In the second step, additional follow-up experiments explore the behavioral mechanism underpinning this result. Third, a structural model integrates the findings into a preference framework. Finally, a vignette study showcases the relevance of the experimental findings in labor market and career choice scenarios.

In the main experiment, participants ("spectators", $n = 653$) judge how much money other people ("workers") deserve for their effort in a piece-work job. The workers initially earn a randomly assigned piece-rate (their circumstances). They know that their piece-rate can either be high (\$0.50) or low (\$0.10) with 50% chance each. Chance determines that one worker receives the high rate, whereas the other worker receives the low rate. Workers decide freely how many tasks they want to complete (their effort choice). Unsurprisingly, workers work much harder and complete roughly three times as many tasks for the higher piece-rate (the endogeneity of choices). In the second step, each spectator is assigned to one pair of workers and informed about their task and circumstances. Spectators decide which final reward each worker deserves. In multiple scenarios, they can redistribute the earnings between the two workers, conditional on workers' effort choices. These merit judgments are the central outcome variable of the study.

The experiment exogenously varies in which circumstances workers make their effort choices. In the *control* condition, the workers do not know their realized piece-rates yet. They only know their odds to obtain 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 and subject to the same situational influence – 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 and advantaged by these circumstances, whereas workers with a low piece-rate are discouraged and disadvantaged. Thus, in the treatment condition, but not in the control condition, the endogeneity of choices differentially (dis)advantages the workers. I compare spectators' merit judgments across the two conditions. Do merit judgments reward the same effort choices equally across conditions, thereby ignoring the external circumstances in which workers make their decisions? Alternatively, do spectators compensate the disadvantaged workers in the treatment condition for the fact that they are discouraged from working hard?

The results show that participants' merit judgments are completely insensitive to the endogeneity of choices. The spectators strongly redistribute payments to reward workers for higher effort, but they do so equally in both conditions. They neglect that the disadvantaged worker is discouraged from working hard in the treatment condition but not in the control condition. The average reward share of the disadvantaged worker is even (insignificantly) 0.49 percentage points (pp) lower in the treatment than in the control condition. The large sample size allows me to rule out even minor increases in the reward of the disadvantaged worker (0.8 pp 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 endogenous and shaped by external situational influence over which the workers have no control.

Why do spectators neglect the endogeneity of workers' choices? To shed light on the behavioral mechanism behind this finding, I run tailored follow-up experiments. I start by investigating whether spectators underestimate the power of situational influence, in line with the well-known *fundamental attribution error* (Ross, 1977). I measure incentivized beliefs about how strongly the piece-rates influence workers' effort choices. However, spectators even slightly overestimate the piece-rate effect, so that its neglect cannot simply be attributed to biased beliefs. Of course, this does not rule out that the endogeneity of choices escapes spectators' *attention* while rewarding the workers. In the second step, I therefore implement an attention intervention ($n = 274$) in which I draw spectators' attention to the effect of situational influence just before their merit judgments. However, merit judgments remain insensitive to the endogeneity of choices even then.

Compensating for disadvantageous situational influence also raises the question of what the two workers to whom a spectator is assigned would have done in identical circumstances. This *counterfactual* is unknown and uncertain even for spectators who accurately anticipate the average piece-rate effect. Therefore, I test for the role of counterfactual reasoning in another experiment ($n = 945$) in which I provide a subset of spectators with accurate information about what the disadvantaged worker would have done in the advantaged environment. I find that, on average, spectators' merit judgments react strongly to the counterfactual effort choice of the disadvantaged worker. Once the counterfactual is revealed to them, they take the endogeneity of choices into account and compensate workers who are disadvantaged by external situational influence. This suggests that the uncertainty of the counterfactual – what would have happened on a level playing field – explains why merit judgments are insensitive to the endogeneity of choices. When the counterfactual is unknown, spectators simply base their merit judgments on the only clear and reliable evidence they have, namely the observed effort choices. This results in a "burden of the doubt" for the disadvantaged worker.

The average results discussed earlier conceal that merit judgments are vastly heterogeneous. In the next step, I therefore estimate a structural model of merit views to assess the prevalence of different merit views in the population. The model builds on a simple theoretical framework that I sketch in the introductory Section 2 and thus brings the study's argument full circle. I distinguish between four distinct merit views: comparable choice meritocrats, actual choice meritocrats, libertarians, and egalitarians. "Comparable choice meritocrats" hold workers accountable for the counterfactual effort choices that workers would make in identical, comparable circumstances, but – in line with the reduced-form results – potentially discount this counterfactual when it is unknown and uncertain. "Actual choice meritocrats" reward workers proportional to their actual effect choices, even if these choices are endogenous to external circumstances. "Libertarians" accept any inequality and do not redistribute. Lastly, "egalitarians" think that the workers always deserve equal payments. The estimated model classifies 26% of participants as comparable choice meritocrats. In line with the reduced-form results, I estimate that they fully neglect situational influence when the counterfactual is uncertain. Meanwhile, 37% of participants are classified as actual choice meritocrats, 23% as libertarians, and 14% as egalitarians. The results show that people hold fundamentally different merit views. Importantly, they also reveal that, even in a (counterfactual) world where counterfactual choices were known, only about 26% of individuals would compensate for disadvantageous situational influence. The prevailing meritocratic fairness ideal ignores the endogeneity of choices.

Although the controlled experimental environment comes with the crucial advantage that the effect of interest is clearly and credibly identified, it also comes at a cost: It differs from many real-life settings that characterize the debate about merit, choices, and circumstances. To mitigate this concern, I run a vignette study (n = 1,222) showing that the insensitivity of merit judgments to the endogeneity of choices can also be observed in labor market and career choice scenarios. For instance, 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 a discouraging environment with few opportunities and incentives to work hard. In both cases, the choice not to work hard legitimizes a highly unequal outcome, irrespective of the disadvantageous external situational influence.

Discussion The pros and cons of meritocracy have been the subject of a heated public debate (Frank, 2016; Greenfield, 2011; Markovits, 2019; Sandel, 2020; Young, 1958). Meritocratic fairness promises that the family, neighborhood, and circumstances one is born into should not matter – a popular notion that closely connects to the prominent ideas of equal opportunity and the American dream. However, the findings of this study suggest that meritocratic fairness is likely to be "*shallow*". Even though meritocratic fairness holds that individuals should not be judged by their external circumstances, people neglect that these external circumstances also influence the choices that agents make and hold them fully responsible for these choices. Thus, choices "launder" unequal circumstances and legitimize the ensuing inequality.

In practice, not only effort but also valuable talents and abilities, such as cognitive skills, are viewed as meritorious and worthy of reward. These talents, skills, and personality traits are also shaped by external circumstances, in particular, during early childhood (e.g., Alan and Ertac, 2018; Heckman, 2006; Kosse et al., 2019; Putnam, 2016). Hence, while this study focuses on the endogeneity of *choices*, an analogous question arises for the endogeneity of *skills*. The former is the starting point of this study because it is the simpler, more transparent, and relatable channel. Because individuals ignore the endogeneity of choices – an effect they should be well familiar with –, I expect that a similar neglect also arises for the endogeneity of *skills*.

Of course, holding others responsible for their actual choices (or skills) may simply be a practical necessity of living together. The results of the study thus connect to an old theme in the philosophy of responsibility (Eshleman, 2016; Nelkin, 2019), but the study neither can nor aims to settle this normative debate. Instead, it documents which merit views people endorse in practice.

These views on fairness matter because they characterize the society in which we live. Ultimately, the neglect of endogeneity is likely to shape which policies voters demand. "Shallow meritocrats" endorse *predistribution* policies that level the playing field and equate circumstances ex-ante. Yet, they are reluctant to compensate others for unequal circumstances via *redistribution* after unequal choices have been made. This could explain why ex-post policies such as affirmative action are considered controversial and suggests that policymakers who want to mobilize support for advancing equality of opportunity should emphasize ex-ante, predistributive policies.

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; Fisman et al., 2020; Giuliano and Spilimbergo, 2013; 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 to work hard or to take 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; Krawczyk, 2010; Mollerstrom et al., 2015). Small differences in merit sometimes justify large reward inequalities (Bartling et al., 2018; Cappelen et al., 2020a). Moreover, Cappelen et al. (2020c) show that even degenerate choices can have 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, merit judgments seem to be all about choice. By contrast, luck and circumstances outside the agents' control are commonly rejected as a legitimate source of merit. However, how do merit judgments deal with the ubiquitous endogeneity of choices to external circumstances? This study is the first to address this question and provide 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; Falk et al., 2020b; Gurdal et al., 2013; Nagel, 1979). Individuals are often approved or disapproved not only for 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 if it directly shapes the decision they make.

This study also connects to a recent literature on inference in economics (e.g., Benjamin, 2019; Enke and Zimmermann, 2017; Graeber, 2021; Han et al., 2020; 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) – a key element of counterfactual reasoning. However, counterfactual reasoning itself remains relatively unexplored in economics, even though cognitive scientists have long since acknowledged its centrality to causal reasoning and inference (Byrne, 2016; Kahneman and Miller, 1986; Lagnado and Gerstenberg, 2017; Roese, 1997; Sloman, 2005). This study illustrates that counterfactual reasoning is a potent mechanism. The inherent uncertainty of the counterfactual strongly affects individuals' choices even though they accurately anticipate the expected counterfactual.

The remainder of the paper is structured as follows. Section 2 sets the stage by discussing a simple conceptual framework of merit views, Section 3 describes the main experimental design, and Section 4 presents the main results. Section 5 examines their behavioral foundations, Section 6 structurally estimates the model of fairness views, and Section 7 reports the vignette study. Finally, Section 8 concludes the paper.

2 Conceptual framework

To fix ideas, I introduce a simple theoretical framework that directly maps into the experimental design. Two workers, $k \in \{A, B\}$, independently choose how much effort E_k they exert, given their external circumstances, namely their returns to effort π_k. As in the experiment, the workers' returns to effort are externally determined by a lottery. Worker A randomly receives a high piece-rate, whereas worker B randomly receives a low piece-rate. The workers have convex effort costs $\frac{1}{2}(E_k - \theta_k)^2$, where θ_k is their diligence or taste for hard work. Hence, worker k maximizes $\pi_k E_k - \frac{1}{2}(E_k - \theta_k)^2$, chooses the optimal effort level

$$E_k^* = \theta_k + \pi_k,$$

and earns $P_k = \pi_k E_k^*$. The optimal choice E_k^* can be decomposed into an "internal" cause (θ_k) and an "external" cause (π_k). Thus, conditional on their types θ, worker A works harder due to their higher returns to effort. Worker A (high piece-rate) is advantaged, whereas worker B (low piece-rate) is disadvantaged by external situational influence.[1]

How is workers' merit in this setting evaluated? Suppose that a neutral third person observes this situation. In line with the literature on fairness preferences, I refer to the third party as "spectator" because the spectator's own monetary payoff is not at stake. The spectator (hereafter referred to as "she" or "her") observes the workers' (referred to as "he" or "him") circumstances, the share of the total payment that the disadvantaged worker B receives $p = \frac{P_B}{P_A+P_B}$, and the share of total work that he conducts $e = \frac{E_B}{E_A+E_B}$. Without loss of generality, I focus on the disadvantaged worker B because he will be at the center of the later analysis. Moreover, I focus on his payment and effort shares (denoted by lower case letters) because they can easily be compared across situations. The spectator can redistribute the workers' earnings to implement the reward share r of worker B that she prefers. Redistribution comes at no cost.[2] I assume that spectator i maximizes the utility function

$$U(r_i) = -\frac{1}{2}\left[r_i - m_i(e,s)\right]^2$$

where $m_i(e,s)$ denotes i's merit view, that is, her view about which reward the disadvantaged worker deserves for providing the effort share e in the external situation s. Thus, the spectator wants to implement the reward share r_i that she thinks is merited by worker B in situation s.

$$r_i^* = m_i(e,s)$$

This set-up combines several features that are well-suited to characterize merit judgments. First, it focuses on choices about effort and hard work that play a major role in the debate about merit. Second, it deals with *relative* merit judgments, that is the merit of worker B compared to worker A. After all, "high" or "low" effort and "high" or "low" rewards can most easily be distinguished in comparison. Third, rewards are assigned via redistribution to mirror the fact that society's fairness views are often implemented via redistributive schemes that intervene into naturally arising market outcomes.

[1] I abstract from income effects on labor provision (i.e., worker's utility is linear in money) because income effects will arguably be absent in the experimental application. The structural assumptions on the effort cost function C are made for illustrative purposes only. The argument in this paper depends mainly on $\partial E_k^*/\partial \pi_k > 0$.

[2] For simplicity, I abstract from the frequently studied fairness-efficiency trade-off. Existing research shows that fairness concerns often dominate efficiency concerns (Almås et al., 2020).

I consider four distinct merit views.

Actual choice meritocrat: $m_i(e,s) = e$
For "actual choice meritocrats", choice is the only relevant criterion of merit. They hold people fully responsible for their choices, even if these choices are endogenous to external circumstances. In the worker setting sketched earlier, actual choice meritocrats hold that the disadvantaged worker B deserves a payment share equal to his effort share. For instance, he deserves 25% of the payment if he completed 25% of the tasks; he deserves 75% of the payment if he completed 75% of the tasks.

Comparable choice meritocrat: $m_i(e,s) = \hat{E}_i c(e,s)$
"Comparable choice meritocrats" do not hold individuals responsible for external causes of choice (π) but only for internal (θ).[3] To subtract any external influence on choice, a comparable choice meritocrat asks, "What would the two workers have done in an identical, comparable situation?" Merit is derived from these counterfactual, comparable effort choices. Accordingly, comparable choice meritocrats think that the disadvantaged worker B deserves a payment share equal to the counterfactual effort share c that he would have provided had he been in the same advantaged circumstances as worker A.[4] Since comparable choice meritocratism requires an inference about the counterfactual, biased counterfactual reasoning could lead to a discrepancy between the perceived counterfactual effort share $\hat{E}_i c(e,s)$ and the actual but unknown counterfactual effort share $c(e,s)$.

Egalitarian: $m_i(e,s) = 50\%$
The workers always deserve equal payment shares (as in Almås et al., 2020).

Libertarian: $m_i(e,s) = p$
Any pre-existing earning share p is regarded as legitimate and accepted (as in Almås et al., 2020).

In sum, actual choice meritocrats equate merit and effort even if external circumstances shape the effort choices. By contrast, comparable choice meritocrats identify merit with counterfactual effort choices in identical, comparable circumstances. Because the counterfactual is uncertain, their merit judgments also depend on their inference and counterfactual reasoning. The other two types, egalitarians and libertarians,

[3]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; Heckman, 2006; Kosse et al., 2019). Ultimately, one could hence even ask whether these differences can justify merit differences. However, this question is outside the scope of this paper.

[4]In principle, comparable choice meritocrats could also base their merit judgments on counterfactual effort choices in another environment, for example, the low piece-rate situation. Relatedly, Roemer (1993) takes an individual's relative ranking in the effort distribution conditional on circumstances, $f(E_k^*|\pi_k)$ as a comparable measure of merit. These details affect neither the qualitative argument here nor the qualitative interpretation of later treatment effects.

do not condition merit on choice. They respectively accept no or any form of unequal rewards and play only a minor role in the context of this study.

Conceptually, there are intriguing *normative* arguments for both actual choice and comparable choice meritocratism. 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.

Here, however, the research question is of *positive* nature: Which merit judgments does the general population make? First, are they sensitive to the endogeneity of choices? Second if not, are they insensitive because comparable choice meritocrats are absent from the population or because they misinfer what would have happened without situational influence and fail to apply their merit view? The main experiment sets out to investigate the first question in an environment that mirrors the simple model sketched above. Tailored mechanism experiments follow to explore the second question.

3 Experimental design

Studying how the endogeneity of 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 the situational influence of external circumstances on choices. Below, I describe how I tailor the experimental design to meet both requirements.

3.1 Setting: Redistribution task

I create an experimentally controlled situation of inequality between *workers* and observe how study participants (*spectators*) redistribute money between the workers, conditional on workers' effort choices. Spectators decide what each worker deserves and thereby judge which merit originates from the workers' choices. The set-up is in line with the framework sketched in Section 2.

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 requires no special qualification but demands effort and concentration, ensuring that hard work determines success rather than skill. Each worker k earns a piece-rate π_k 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). A worker's initial payment 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 full instructions for the 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 workers A and 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 C), rendering the setting well-suited to study how merit judgments react to situational influence.

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 informed about the workers' task, situation, choices, and earnings. 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.[6] Redistribution comes at no cost. Spectators know that their decision is strictly anonymous and that workers are unaware of the redistribution stage. Appendix E provides the main instructions for spectators, and the full 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; Cassar and Klein, 2019; Mollerstrom et al., 2015). They mirror the fact that society's fairness views are

[5] In the experiment, I randomly vary whether worker A or worker B is the worker with the advantageous, high piece-rate. Here, I recode all responses as if worker A was the advantaged worker to ease analysis and exposition. Reassuringly, Table B.5 shows that spectators' redistributive behavior is insensitive to whether worker A or worker B is advantaged. Moreover, sometimes both workers of 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 experiments that I will introduce later.

[6] Spectators cannot redistribute the fixed remuneration of $1 but only the performance-based rewards.

often implemented via redistributive schemes that intervene into naturally arising market outcomes – a feature that I want to capture in the experiment. I implement the merit judgments of 100 randomly selected spectators so that spectators' decisions are (probabilistically) incentivized. After all, their decisions can have real and meaningful consequences for the workers.[7]

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 hypothetical, 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, whereas worker B completes no task at all ($e = 0\%$). In Scenario 4, both workers complete 25 tasks ($e = 50\%$). Moreover, 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 workers' final payoff. However, spectators are not told which scenario is real and hence have to take each of their decisions seriously.[8] Effort choices in the real scenario vary across experimental conditions (introduced in the following) due to the incentive effects of the conditions. Thus, the real scenario does not allow a consistent comparison across treatments. To circumvent this problem, I only analyze the merit judgments in the first seven scenarios. The contingent response method is central 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 Experimental conditions: Varying situational influence

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

[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.

[8]Indeed, only a few spectators can distinguish the hypothetical scenarios from the real one, even after they saw all scenarios and made all of their redistribution decisions. When I ask them to guess which of the scenarios is real, 46% respond that they do not know. Among the others, only 16% guess correctly. Thus, the recognition rate is only slightly higher than what would be expected under random guessing (12.5%). The results are robust to excluding respondents who recognize the real scenario (see Appendix B.2).

Table 1 Overview of effort scenarios, experimental conditions, and studies

(A) Effort scenarios (presented in random order)

Effort share of worker B: e	(1) **0%**	(2) **10%**	(3) **30%**	(4) **50%**	(5) **70%**	(6) **90%**	(7) **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%)

Contingent response method: 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.

(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 with 50% each	\$0.50 or \$0.10 with 50% each	\$0.50	\$0.10

(C) All experimental studies (for later reference)

Study	**Section**	**Description**
Main study	3, 4	Varies whether endogeneity of choices (dis)advantages workers.
Attention study	5.2	Shifts attention towards endogeneity of choices.
Counterfactual study	5.3, 6	Reveals what would have happened in equal circumstances.
Vignette study	7	Explores merit judgments in exemplary real-world scenarios.
Robustness		
"Equal rates" study	4	Replicates main study, but workers receive same piece-rate.
"Disappointment" study	4	Explores motive to compensate workers for disappointment.
"Equal rates" attention study	5.2	"Equal rates" version of the attention study (see above).

Notes: Panel A presents an overview of all effort scenarios. Panel B summarizes and compares the experimental conditions. Panel C lists all experimental studies that I present in this paper. Only the main study is introduced in this section. The details of all other studies will be introduced in later sections.

Control: 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 completing their work.

Treatment: 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 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 taste for hard work. In the treatment condition, the workers face different circumstances and their effort choices are differentially affected by situational influence. 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 may reflect his higher taste for work or the advantageous situational influence. Do spectators account for this? By comparing spectators' redistributive behavior across treatment and control, I test whether and how the endogeneity of choices shapes merit judgments.

The contingent response method allows me to study merit judgments and their sensitivity to situational influence 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 taste for hard work 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 conditions 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 disadvantaged by the endogeneity of choices and would work harder for a high piece-rate should compensate him with a higher reward share.

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.

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, the workers complete 12 tasks and earn about $5.40, but both figures vary across experimental conditions. I form 100 pairs with 200 of those workers and use them to incentivize spectators' redistribution decisions.[9]

Spectators I recruit a sample of 653 participants in collaboration with Lucid, an online panel provider which is frequently used in social science research (Coppock and McClellan, 2019, Haaland et al., forthcoming). 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 pre-registered (see Appendix D). The participants are broadly representative of the US adult population 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 follows the characteristics of the American population closely, except perhaps for education: 43% of the sample

[9] I ran the main experimental conditions together with robustness and mechanism experiments 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.

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.2 shows that the covariates are balanced across experimental conditions.

The experiment took place 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, the participants answer a series of demographic questions, which 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 guarantees 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, I ask a series of follow-up questions to collect additional demographic variables and probe for possible mechanisms.

3.4 Additional experiments

I run a series of additional experiments to explore the robustness of the results and shed light on its behavioral mechanisms. The details will be introduced in later sections. For later reference, Panel C of Table 1 provides an overview and brief description of all experiments.

4 Main result

I start by studying spectators' merit judgments in the control treatment. Here, workers' effort choices are comparable because they are made 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 share of the total earnings that a spectator assigns to the disadvantaged worker. 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). In fact, many participants reward worker B proportionally to his effort share. They implement the payment shares that would have occurred if both workers had earned an identical rate (Appendix Figure B.1). Thus, in the control condition where both workers react to the same environment, merit derives mostly from effort choices.[10]

This sets the stage for my main research question. Do spectators take the endogeneity of effort choices into account? In the treatment condition, workers learn about their realized piece-rates already before they make their effort choice. Consequently, worker B is disadvantaged as he endogenously reacts to a discouragingly low piece-rate of $0.10. By contrast, worker A is encouraged by a high piece-rate of $0.50. Do spectators assign a higher reward share to worker B in the treatment than in the control condition to compensate him for this disadvantageous situational influence?

The results show that merit judgments are fully insensitive to the endogeneity of choices. Figure 1 shows that the payment shares are 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). Hence, spectators do not compensate worker B for the disadvantageous situational influence in the treatment condition. They even assign an (insignificant) 0.49 pp lower share to him ($p = 0.464$; see Table 3). Panel B shows that this conclusion holds for all seven scenarios. Whether worker A or B completes more tasks, or both work equally hard, spectators do not counterbalance the effect of external situational influence. 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 ($p = 0.668$).[11]

This null result does not reflect a noisy estimate but rather constitutes a precisely estimated null finding. Averaged across scenarios, 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

[10]Deviations from effort-proportional rewards indicate traces of libertarian and egalitarian redistributive behavior. For instance, in effort Scenario 4 where worker B contributes exactly half of the tasks, worker B receives a mean payment share of 40.5% rather than an equal 50.0% share. This is due to “libertarian” spectators who never redistribute and always accept the pre-existing reward share of 17% (see Figure B.1). By contrast, in effort Scenario 1 where worker B completes no task at all, he still receives an average reward share of 7.8%. This is due to “egalitarian” spectators who always implement equal shares irrespective of the workers’ effort decisions (see Figure B.1).

[11]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 on the participant level.

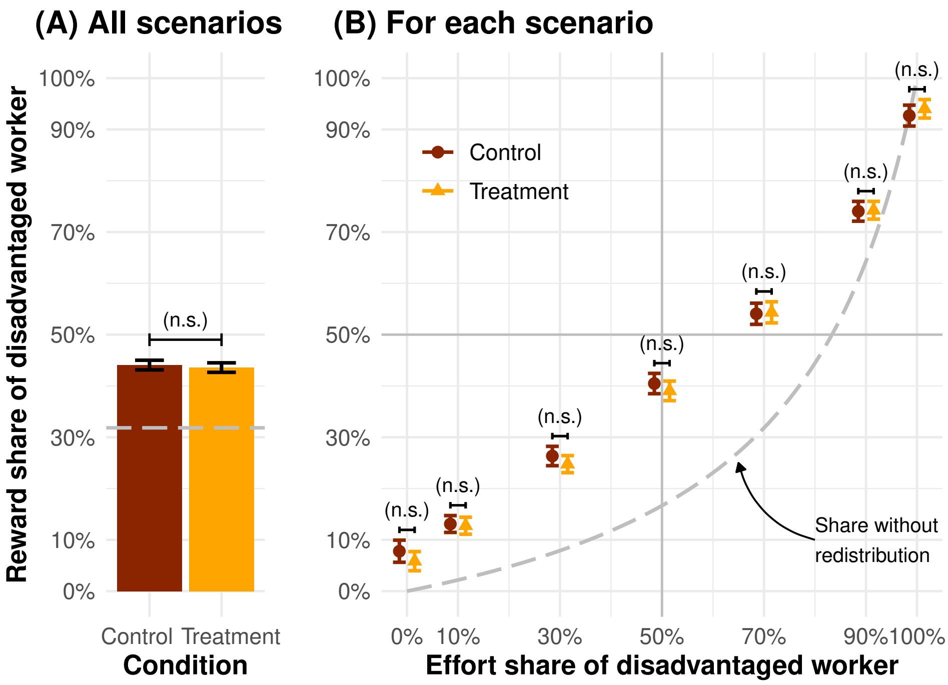


Figure 1 Average reward share of disadvantaged worker with 95% CI

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.) $p \geq 0.10$.

situational influence receive a compensation of more than 0.8 pp of the total payment. The results thus provide strong evidence for the absence of a meaningful effect.[12]

An average null effect might still conceal meaningful treatment effects for parts of the population. I therefore test for heterogeneous treatment effects. In the first step, I test for heterogeneity alongside six pre-registered 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 action and abilities instead of attributing them to luck, fate, or the actions of others. None of these variables significantly moderates the treatment effect (see Table B.3).[13] In the 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

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

[13]Moreover, none of the variables is significantly associated with merit judgments in the baseline control condition.

Table 3 Treatment effects on 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). *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

average treatment effect. No significant heterogeneity in treatment effects is detected ($p = 0.446$), which corroborates my main result.

> **Result**: Individual merit judgments fully neglect the endogeneity of choices. People reward others for their effort, even if effort decisions are endogenous to external circumstances.

Robustness

I replicate the results in multiple robustness checks. 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 3, 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).

Second, one might be concerned that the direct effect of the piece-rates on earnings is too salient and crowds out attention to situational influence. 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 thereby overlook that the worker would also have worked much harder (e.g., complete 35 tasks for a payment of \$17.50). In other words, spectators might primarily focus on the fact that, for the same effort choice, the disadvantaged worker would have earned more with a higher piece-rate and simply forget that a higher piece-rate would also have changed his effort choice. However, evidence from an additional experiment (**robustness study: equal rates**, $n = 661$) does not support this explanation. The experiment relies on a between-subject treatment-control variation which is analogous to the main study but keeps the realized piece-rate of both workers constant.[14] As before, both workers have identical expectations about their piece-rate (\$0.10 or \$0.50 with an equal chance) in the control condition. In the treatment 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 by situational influence 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.[15] Consequently, there is no direct piece-rate effect on payments that could distract spectators. Turning to the results, I detect no significant difference in merit judgments across the two conditions. Spectators accept that earnings move proportionally with effort in both conditions. They reward effort – irrespective of whether or not it is shaped by situational influence. This independent robustness experiment thus fully replicates the main results. Again, the null result is obtained with high precision. The 95% confidence interval rules out even small treatment effects (above 0.9 pp), and I observe a null effect in each of the seven effort scenarios (Table B.1, Panel B).

A third 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 larger disappointment. For instance, workers who completed ten tasks hoping for a \$0.50 piece-rate might be more disappointed to learn that they earn only \$0.10 per task (control condi-

[14] I ran the "equal rates" experiment in parallel to the main study in June 2020. The study protocol closely follows the main experiment. As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix A). The results are robust to excluding potentially inattentive responses (misunderstanding of the instructions, survey-taking fatigue, and "speeders"; see Appendix B.2).

[15] Workers who receive a \$0.90 piece-rate are not used for this robustness study and receive their payments without a redistribution stage. Workers with a \$0.10 piece-rate are used in a second "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 in the main text (see Appendix B.2).

tion) than workers who learn about their $0.10 piece-rate already before they complete the ten tasks (treatment condition). 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. Admittedly, the explanation seems unlikely because, to account for the results, the two effects would need to offset each other in all seven effort scenarios. Nevertheless, I design an additional experiment (**robustness study: disappointment**, $n = 606$) that rules out this confounding channel.[16] I replicate the main design with one crucial exception: Workers do not make a choice. Instead, all workers have to complete exactly ten tasks. Since no choice is involved, choices are not endogenous, situational influence on effort choices does not exist, 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 treatment difference (Table B.4). On average, spectators assign a 2.2 pp higher reward share to worker B in the control than in the treatment condition – an effect size that could not even conceal a minor treatment effect.

Lastly, one could argue that the spectators attempt to draw inferences about the workers' life situations outside the experiment. For instance, a worker who completes 25 tasks for a $0.10 piece-rate (treatment) might not only be more diligent than a worker who completes the same amount of tasks for better piece-rate prospects of either $0.10 or $0.50 (control). He might also assign a higher marginal value to money or have lower marginal opportunity costs of time. Spectators could interpret this as a sign of neediness and assign a higher payment share to the disadvantaged worker B in treatment than control. Any such argument predicts the existence of a treatment effect and is thus firmly rejected by the null result.

5 Mechanism

This section investigates why individuals' merit judgments are insensitive to the endogeneity of effort choices. The theoretical framework of Section 2 suggests two explanations. On the one hand, the endogeneity of 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 situational influence ("comparable choice meritocratism"), but they struggle to do so because they fail to infer what would have

[16] 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.4).

happened in identical, comparable circumstances. Here, I explore three behavioral obstacles that could impair spectators' inference: the fundamental attribution error, a lack of attention, and the uncertainty of the counterfactual.[17]

5.1 Fundamental attribution error

Do spectators understand that circumstances affect choices, that is, that workers' effort strongly react to the piece-rate workers earn? It may well be the case that spectators overly attribute choices to the decision-maker and underestimate the role of circumstances. Such an inferential error would be in line with the so-called fundamental attribution error, namely the notion that individuals underestimate situational influence on human decisions (Ross, 1977). 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 out of ten participants earns a $5 Amazon gift card if her response is at most one task away from the true value.

The findings do not support that a fundamental attribution error is driving the neglect of situational influence. Participants believe that workers complete 3.46 times as many tasks for a rate of $0.50 than for a rate $0.10. Thus, the perceived incentive effect is even slightly larger (though not significantly so) than the observed effect of 3.33 (p = 0.749, t-test).

5.2 Attention

Are spectators aware of the endogeneity of effort choices while making their merit judgments? Once asked explicitly about it, participants acknowledge that situational influence exists, but it might still escape their attention while they make their merit judgments. Attention (or a lack thereof) is a powerful explanation of behavior in many other domains (e.g., Andre et al., 2021; Chetty et al., 2009; Gabaix, 2019; Taubinsky and Rees-Jones, 2018). To test for this mechanism, I ran an additional experimental condition that draws participants' attention to the endogeneity of effort choices just

[17]Cappelen et al. (2019) also study fairness views in an uncertain environment but their mechanism can only play a negligible role in my setting. They show that individuals do not want to risk rewarding the wrong person and hence prefer more equal rewards when it is unclear who merits the higher reward. However, in my setting, it is clear for comparable choice meritocrats that worker B merits a (weakly) higher reward in the treatment than in the control condition. It remains only unclear how much higher the reward should be. "Risk-averse" comparable choice meritocrats would still want to compensate the disadvantaged worker when the counterfactual is uncertain to ensure their reward decision is close to the expected fair merit judgment.

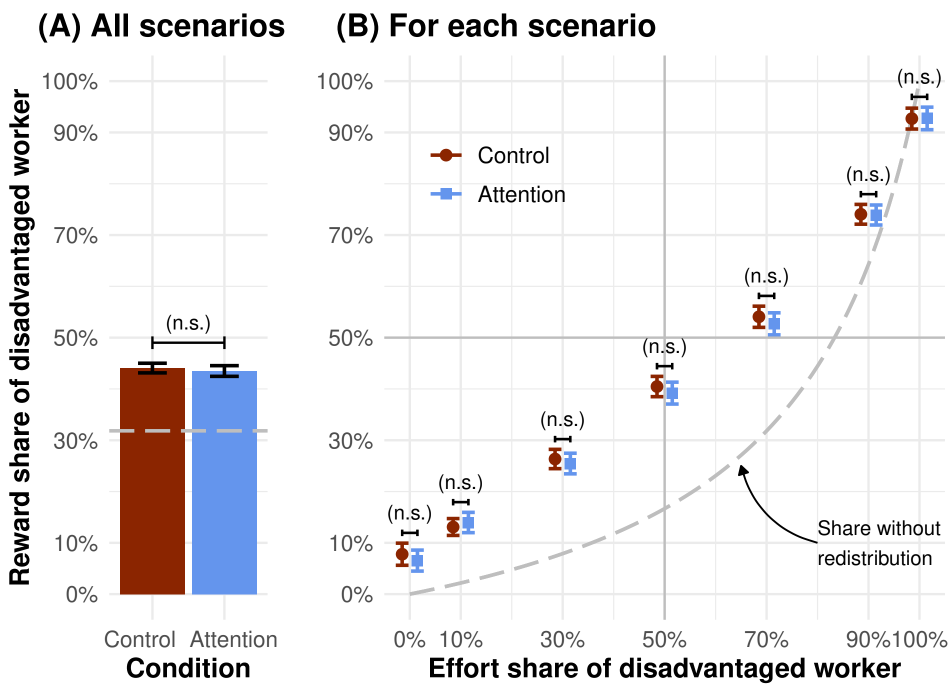


Figure 2 Attention study: Average reward share of disadv. worker with 95% CI

Notes: Results from the attention 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.) $p \geq 0.10$.

before their merit judgments ($n = 274$).[18]

Attention: 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 situational influence is salient to spectators while making their merit judgments. If a lack of attention to situational influence explains its neglect, spectators should compensate the disadvantaged worker with a higher reward share in the attention condition compared to the baseline control condition.

[18] I ran this experiment in parallel to the main conditions in June 2020. The study protocol closely follows the main experiment. As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix A).

This is not the case. Participants who are informed about and focused on situational influence still do not compensate the disadvantaged workers. As before, the null effect is precisely estimated and present in each of the seven effort scenarios (see Figure 2). Aggregated across scenarios, the mean payment share of worker B is 43.5% in the attention condition versus 44.1% in the control condition. The 95% interval of their difference allows me to rule out even tiny treatment effects of 0.8 pp (see Table B.1, Panel C).[19] Hence, a lack of attention to the endogeneity of effort choices also does not explain the results.

5.3 Uncertainty of the counterfactual

Compensating worker B for the disadvantageous situational influence he is exposed to does not only require an understanding and an awareness of the average piece-rate effect. It also raises the concrete question of what the two workers to whom a spectator has been assigned would have done in 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 external situational influence cannot justify merit and hence would want to correct for it.[20] However, this counterfactual is unknown and uncertain even for spectators who accurately anticipate the average piece-rate effect. Recent research shows that people struggle with complex decisions in uncertain and contingent environments, rendering this a promising explanation for why spectators' merit judgments neglect the endogeneity of choices (Esponda and Vespa, 2019; Martínez-Marquina et al., 2019).

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, $n = 945$).[21] 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-rates, are then randomly assigned to one piece-rate, and subsequently have to follow-up on their commitment. Importantly, this technique measures worker's counterfactual effort choice in

[19]The results are robust to excluding potentially inattentive responses (misunderstanding of the instructions, survey-taking fatigue, "speeders"; see Appendix B.2). I also replicate the results in an analogous extension of the robustness experiment with equal piece-rates (**attention: equal rates**, $n = 267$, see Table B.1, Panel D).

[20]As discussed in Section 2, this benchmark is not unique. For instance, a comparable choice meritocrat might also ask what both workers would have done for a low piece-rate of $0.10 or in another common piece-rate environment.

[21]I ran this experiment in January 2021. The study protocol closely follows the main experiment. As before, the sample broadly represents the US population, and treatment assignment is balanced across covariates (see Appendix A). The results are robust to excluding potentially inattentive responses (misunderstanding of the instructions, survey-taking fatigue, "speeders"; see Appendix B.2).

an incentivized way. Thus, I know how many tasks the workers (would) complete for both piece-rates. Spectators are informed about this procedure. As before, they make merit judgments in eight scenarios of which seven are hypothetical and allow me to freely vary the counterfactual effort choice of worker B (contingent response method). Spectators do not know which of the eight scenarios is real so that all of 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 ($e = 30\%$).[22] The next four scenarios are randomly generated and will be used in Section 6. 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 4 provides an overview of all effort scenarios and experimental conditions.

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

Low counterfactual (short: Low): Spectators are informed about worker B's counterfactual effort choice for a high piece-rate. In the "low counterfactual" condition, 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 situational influence.

High counterfactual (short: High): This condition provides information about worker B's counterfactual effort choice, too. Here, however, worker B would complete as many tasks as worker A for a high piece-rate. Situational influence thus exists and strongly affects worker B's choice. Workers A and B (would) make the same choices in the advantaged environment; hence, this information also implies that they share the same taste for hard work.

Figure 3 presents the results (see also Table B.2). First, it reveals that the average reward for worker B is very similar in the "no information" condition and the "low counterfactual" condition.[23] Thus, in the baseline condition with unknown counterfactual, spectators reward worker B as if they knew that his counterfactual effort choice would be no different. This suggests that spectators in the baseline condition base their merit judgments on the assumption that choices have not been shaped by situational

[22] 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 therefore not included.

[23] If at all, spectators are even slightly more generous toward worker B in the "low counterfactual" condition. This difference is significant in the scenario where worker B has an effort share of 30%.

Table 4 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 equals $E_B + X$ where X ranges from 0 to 25.
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 6 to structurally estimate a model of merit views.

influence. They focus on observable effort choices, the only reliable evidence they have, akin to a "burden of the doubt" for the disadvantaged worker.

Second, a comparison of the "low counterfactual" and "high counterfactual" conditions exposes that, once known, the counterfactual choice of worker B matters substantially for spectators' merit judgments. 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. This effect is driven by a subset of spectators who distribute the payment equally once they know that both workers would have worked equally hard for a high piece-rate. About 32% of spectators implement equality in the "high counterfactual" condition, whereas only 7% do so in the "low counterfactual" and "no information" condition respectively (see also Figure B.2).[24]

In short, spectators care about the counterfactual effort choice of worker B. Once known, their merit judgments take situational influence into account and compensate workers who are disadvantaged by external circumstances. This effect is driven by about one-quarter of participants, whereas the remaining participants do not adjust their reward behavior to the counterfactual information. However, *all* participants fully neglect the effect of situational influence when no information on the counterfactual choice is provided. This suggests that, in the presence of an unknown, uncertain counterfactual, spectators base their merit judgments on the only clear and reliable evidence they have,

[24]Could the large effect of the "high counterfactual" treatment be partially driven by an experimenter demand effect? Respondents might interpret the counterfactual information as a hint from the experimenter to make use of the information. However, the null result in the attention experiment renders such an explanation unlikely. Here, the scope for demand effects seems to be higher. Respondents receive two pages of information which strongly emphasize the endogeneity of choices. 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.

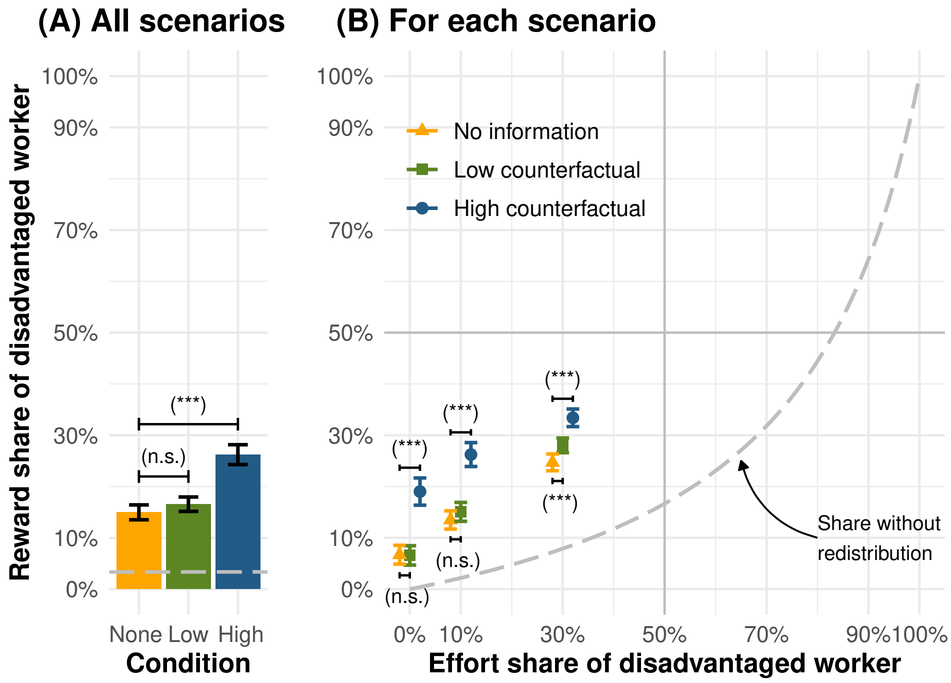


Figure 3 Counterfactual study: Avg. reward share of disadv. worker with 95% CI

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.) $p \geq 0.10$.

namely observed effort choices.

> **Result**: Once the counterfactual is revealed, spectators on average compensate workers for disadvantageous situational influence. The uncertainty of the counterfactual state is thus responsible for the main finding that merit judgments neglect the endogeneity of effort choices.

In light of the model discussed in Section 2, this means that comparable choice meritocrats exist but do not apply their merit view when the counterfactual effort choice under equal circumstances is uncertain and unknown. The next section organizes this and other reduced-form findings in a structured framework.

6 A structural model of heterogeneous merit views

Data from all experiments reveal that individuals endorse distinct fairness types. Typically, the distribution of merit judgments exhibits discrete spikes that coincide with the model of merit views introduced in Section 2 (see Figures B.1 and B.2). In this section, I structurally estimate the model to gauge the prevalence of these fairness views in the population.

6.1 Model and estimation

I assume that each participant rewards the disadvantaged worker B according to $m_i(e,s) + \varepsilon_{is}$. $m_i(e,s)$ is her merit view conditional on worker B's effort share e in situation s, and $\varepsilon_{is} \sim_{iid} N(0,\sigma^2)$ is a normally distributed response error. The model assumes that the population is separated into four distinct fairness types.

Actual choice meritocrats reward workers based on actual effort shares, $m_i(e,s) = e$, irrespective of whether effort choices are endogenous to external situational influence.

Comparable choice meritocrats reward workers based on (counterfactual) effort shares under equally advantaged, comparable circumstances, $m_i(e,s) = \hat{E}_i c(e,s)$, and thus compensate for situational influence. When the counterfactual $c(e,s)$ is known and revealed to the spectators, we have $\hat{E}_i c(e,s) = c(e,s)$. When the counterfactual is uncertain, I assume that comparable choice meritocrats accurately anticipate the expected counterfactual effort share $Ec(e,s)$ but "discount" it and put more weight on the observed effort share e.

$$\hat{E}_i c(e,s) = \rho Ec(e,s) + (1-\rho)e \quad \text{where } 0 \leq \rho \leq 1$$

Both assumptions are in line with the reduced-form results. The discounting of the expected counterfactual could be interpreted as a probabilistic failure to engage in counterfactual reasoning (with probability $1-\rho$) or has a preference to base merit judgments on verifiable information (with weighting factor $1-\rho$).[25]

Egalitarians always implement equality: $m_i(e,s) = 50\%$.

Libertarians fully accept any pre-existing inequality p: $m_i(e,s) = p$.

I use the merit judgments made in Scenarios 4 to 7 of the counterfactual study to estimate the model. These scenarios randomly vary the effort share of both workers and, in the counterfactual conditions, the counterfactual effort share of worker B (see Table 4, Scenarios 4-7). They cover a rich variety of cases and are hence ideally suited to estimate how common different merit views are. Moreover, this procedure allows me to explore the replicability of my reduced-form findings, which do not depend on data from Scenarios 4 to 7. I estimate six parameters, namely the population shares of each

[25] I calibrate $Ec(e,s)$ to the worker data. Appendix B.4 shows that the results of the model are insensitive to two different calibration approaches.

preference type together with the discount parameter ρ and the standard deviation of the response error σ.

The parameters are identified by the within-subject variation in effort scenarios and the between-subject variation in experimental conditions. For example, the share of egalitarians is reflected in the number of individuals who equalize payments in all effort scenarios. Likewise, the share of comparable choice meritocrats becomes evident in the conditions where the counterfactual is known. Here, the share influences how many respondents are willing to redistribute payments according to counterfactual effort shares. In turn, the discount parameter ρ can be identified in the condition where the counterfactual is uncertain and the merit judgments of comparable choice meritocrats crucially hinge on the discounting of the expected counterfactual.

I employ a constrained maximum likelihood procedure. Appendix B.4 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 or an exclusion of participants who initially failed a control question. I also confirm the numerical stability of the maximum likelihood estimator in Monte Carlo experiments.

6.2 Results

The model estimates that 37% of the population are actual choice meritocrats, while 26% are comparable choice meritocrats. Libertarians and egalitarians have a population share of 23% and 14%, respectively (see Table 5). Thus, a large majority of participants, namely 63%, endorse a meritocratic fairness ideal.[26] However, most meritocrats are actual choice meritocrats (about 60% of all meritocrats). They ignore that workers' choices are shaped by unequal situational influence, even if they know what would have happened in equal circumstances. Only a few individuals are comparable choice meritocrats and prefer to take the endogeneity of choices into account. For them, I estimate a ρ of 0.00 which means that even they fully discount counterfactual choices if the counterfactual is uncertain.[27]

The estimated model mirrors the reduced-form results. For instance, a ρ of 0.00 explains why merit judgments are entirely insensitive to situational influence in the conditions where the counterfactual is unknown. Likewise, the model estimates a share of comparable choice meritocrats of 26% which aligns with the observation that a quarter

[26]The estimated share of meritocrats is much higher than in Almås et al. (2020) who classify 37.5% of the US population as meritocrats. In their setting, spectators receive only coarse, binary information about effort choices, namely which of two workers is more productive. Merit presumably plays an even larger role in my setting because the piece-rate task provides a clear and fine-grained measure of effort.

[27]The estimate for ρ is on the boundary. Standard inference in constrained maximum likelihood models can become unreliable if one of the parameters is on or near the boundary (Schoenberg, 1997). In Appendix B.4, I run simulation experiments to show that the inference is nevertheless reliable.

Table 5 Results of the structural estimation

	Estimate	**95% confidence interval**
Population shares		
Actual choice meritocrats	36.7%	[33.0% – 40.3%]
Comparable choice meritocrats	26.2%	[22.8% – 29.6%]
Libertarians	23.0%	[20.2% – 25.7%]
Egalitarians	14.2%	–
Counterfactual discount parameter		
ρ	0.00	[0.00 – 0.09]
Error term and sample		
σ noise	9.27	[9.06 – 9.49]
Respondents	945	
Decisions	3777	

Notes: Results from the counterfactual study, decisions 4-7, maximum likelihood estimation of the structural model of merit views. The estimates indicate the population shares of different fairness views and the uncertainty discount parameter ρ. No confidence interval is reported for the share of egalitarians because their share is deduced from the other estimates. See Appendix B.4 for further details.

of respondents is responsible for the treatment effect in the counterfactual experiment (see Section 5.3). To give another example, the estimated libertarian share of 23% is broadly consistent with the fact that, depending on the effort scenario, 18% to 29% of respondents accept the pre-existing inequality (see Figures B.1 and B.2).

Does the composition of fairness types or the uncertainty discount parameter ρ vary across different parts of the population? To answer this question, I re-estimate the model and allow its parameters to vary across two separate groups of the population (see Appendix B.4). I compare female versus male respondents, above-median versus below-median respondents, respondents with versus without a college degree, and Republicans versus Democrats. I detect no significant differences across groups. In particular, I estimate a ρ of 0.00 in each group, which suggests that the neglect of uncertain counterfactual states is a fundamental feature of merit judgments (see Table B.7).

Taken together, the results show that people endorse fundamentally different merit views. Crucially, even if the counterfactual choice were known (arguably a rare if not "counterfactual" situation in the real world), only about 26% of individuals would compensate for situational influence. Thus, the prevailing fairness ideals ignore the endogeneity of choices.

Result: A structural model of merit views classifies only 26% of individuals as comparable choice meritocrats who want to correct for the endogeneity of choices. Replicating earlier results, the model also estimates that even comparable choice meritocrats fully neglect the endogeneity of choices when the counterfactual is uncertain.

7 Vignette study with real-world scenarios

The controlled set-up of the online experiment has many advantages. In particular, it measures merit judgments in situations with real consequences, and it allows for an exogenous variation of external situational influence. However, its stylized environment – two crowd-workers, working for a randomly assigned piece-rate, earning up to $25 – 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 endogeneity of choices in three real-world scenarios. I report results from an additional vignette study ($n = 1{,}222$) which sheds light on the following three questions, chosen as common and important practical examples of merit judgments: Are minorities compensated for the detrimental choices they might make because they are discriminated? Is a person growing up with few opportunities and incentives to exert effort blamed for being idle? And 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? The study was run in February 2021 in collaboration with the survey company Lucid. Respondents were recruited from the general US population.[28]

7.1 Vignettes

Each vignette describes a simple hypothetical scenario with two people that are exposed to unequal situational influence. The person disadvantaged by situational influence earns much less money due to the detrimental choice he makes. Below, I outline each vignette.[29]

> **Discrimination vignette**: A white and a black employee compete for a promotion which comes with a one-time bonus of $10,000. However, their boss is notorious for being racist, and he never promotes black employees. The white employee works hard to win the promotion, the black person does not, and the white employee is promoted.

[28]The study was conducted in two waves. Wave 1 was collected together with the *robustness study: disappointment*. Here, every respondent faced two randomly selected vignettes. Wave 2 was launched shortly thereafter, and respondents faced all vignettes in random order. I exclude respondents who speed through the survey and complete the vignettes with an average response time of less than one minute. The results are robust to both stricter and more lenient exclusion criteria (see Table B.8). Table A.1 shows that the sample does not fully match the characteristics of the general population. Among others, the sample contains more females, 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 Appendix B.5).

[29]The full wording of the vignettes is presented in Appendix F. 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 Appendix B.5).

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.

Start-up vignette: The 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.

Analogous to the main experiment, respondents can specify how much money each person deserves by hypothetically redistributing the income between the two people. If their merit judgments are sensitive to situational influence, they should compensate the disadvantaged person for the adverse situational influence that shaped his choice. Redistribution toward the disadvantaged person could, however, 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 start-up vignette.

To identify the sensitivity of merit judgments to situational influence, 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: The vignettes describe only the actual decisions of both persons.

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

Table 6 Merit judgments in the vignette study

(A) Share of respondents redistributing towards the disadvantaged worker

	Binary indicator for compensation			
	Discrimination	Poverty	Start-up	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	Start-up	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 on 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 toward 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. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

7.2 Results

Table 6 summarizes the results. Once more, I find that merit judgments neglect the endogeneity of effort choices. First, the neglect of situational influence already induces little redistribution toward the disadvantaged person in the baseline condition. For instance, in the discrimination vignette, only 42% of 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 pay-off (Column 2, 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 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, situational influence is present (though uncertain), whereas it is verifiably absent in the "low counterfactual" condition. If, as in the main experiment, baseline merit judgments are insensitive to situational influence, they should be similar across the baseline and the "low counterfactual" condition. Indeed, the reward decisions are virtually identical in both conditions. Pooled across vignettes, only 0.4 pp more respondents redistribute money toward the disadvantaged person in baseline than in "low counterfactual" (Column 4, Panel A). Likewise, the average reward share of the disadvantaged 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 toward 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, whereas they are more muted in the start-up 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 only integrate situational influence in their merit judgments once the counterfactual is known but ignore it if the counterfactual is uncertain.

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

Result: Merit judgments neglect the endogeneity of choices also in important real-world scenarios.

8 Concluding remarks

The idea of meritocracy has become central in Western politics where it has shaped the public debate, the economic culture, and social reforms. Meritocracy promises that the family, neighborhood, and circumstances one is born into should not matter. This promise is popular and closely connects to the prominent ideas of equal opportunity and the American dream. However, the findings of this study suggest that, in practice, meritocratic fairness is likely to be "*shallow*". Even though it claims that individuals should not be judged by their external circumstances, people ignore that these external circumstances also influence the choices that agents make.

In a series of experiments with about 4,000 participants from the general US population, I document that individuals reward and penalize workers for their effort choices, even if their choices are strongly endogenous to and shaped by external circumstances. I experimentally identify the uncertainty of the counterfactual – what the disadvantaged person would have done in advantaged circumstances – as the cause of the neglect. Only once the uncertainty of the counterfactual is resolved and participants know what would have happened on a level playing field, about a quarter of respondents start to compensate for the disadvantageous endogeneity.

The uncertainty of the counterfactual state is often an inevitable feature of reality, and so is, this suggests, the neglect of endogeneity. Therefore, it seems likely that the neglect is common also outside the US and extends to other determinants of merit, such as cognitive skills, personality traits, or educational achievements, which are also highly endogenous to and shaped by circumstances (e.g., Alan and Ertac, 2018; Heckman, 2006; Kosse et al., 2019; Putnam, 2016).

A structurally estimated model of merit views reveals that the prevailing fairness ideal ignores the endogeneity of choices, even when the counterfactual state is known. Most participants endorse *actual choice meritocratism*: They reward and hold workers responsible for their observable effort choices, irrespective of whether choices are endogenous to external circumstances.

Of course, holding others responsible for their actual choices may simply be a practical necessity of living together. Any fairness principle must also be evaluated in terms of its prospective incentive effects. Actual choice meritocratism provides clear guidance to both agents and spectators. By contrast, comparable choice meritocratism could create a complicated signaling game where disadvantaged agents try to signal high counterfactual effort choices strategically, while spectators anticipate this behavior and face even greater difficulties in inferring the counterfactual. This may explain why actual choice meritocratism is more popular in the US population and why even comparable choice meritocrats account for the endogeneity of choices only if they have access to reliable information about the counterfactual state.

The structure of merit judgments is likely to affect which policies voters demand. In particular, "shallow meritocrats" may accept the consequences of unequal opportunities, even though they oppose unequal opportunities themselves. Once unequal opportunities led to unequally meritorious choices, these choices can justify the resulting inequality. Consequently, meritocrats endorse *predistribution* policies that level the playing field and equate circumstances ex-ante. By contrast, they are more reluctant to compensate others for unequal circumstances via *redistribution* after unequal choices have been made. In practice, a policymaker is therefore likely to face much larger support for predistributive than for redistributive policies. This may also explain why many affirmative action policies are considered controversial and often depicted as undermining the merit principle (Harrison et al., 2006), even though they attempt to correct for the unequal opportunities that agents faced in producing merit. Simply put, in a meritocracy, choices can launder circumstances and legitimize the ensuing inequality.

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.

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, *qjab006*.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective Models of the Macroeconomy: Evidence From Experts and a Representative Samples," *Working Paper*, 2021.

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.

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.

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.

_ , Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden, "A Meritocratic Origin of Egalitarian Behavior," *Working Paper*, 2019.

_ , 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*, 2020, (forthcoming).

Carrell, Scott E., Marianne E. Page, and James E. West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Cassar, Lea and Arnd H. Klein, "A Matter of Perspective: How Failure Shapes Distributive Preferences," *Management Science*, 2019, *65* (11), 4951–5448.

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.

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

Eshleman, Andrew, "Moral Responsibility," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, 2016.

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.

__ **, 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.

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*, 2021.

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*, (forthcoming).

Han, Yi, Yiming Liu, and George Loewenstein, "Correspondence Bias," *Working Paper*, 2020.

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.

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.

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

Nelkin, Dana K., "Moral Luck," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, 2019.

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.

Schoenberg, Ronald, "Constrained Maximum Likelihood," *Computational Economics*, 1997, *10* (3), 251–266.

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?," *Working Paper*, 2021.

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.

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

Appendices

Table of Contents

A Samples

Overview The table provides an overview of all spectator samples used in this study. It lists all samples and describes when and how they were collected.

Sample	When	How	Population	Recruitment	n
Main study	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	653
Robustness study "Equal rates"	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	661
Robustness study "Disappointment"	February 2021	Online experiment	US adults	Via survey company Lucid	606
Attention study	June 2020	Online experiment	US adults (targeted*)	Via survey company Lucid	274
Attention robustness study "Equal rates"	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
Vignette study	February 2021	Online survey	US adults	Via survey company Lucid	1,222**
				Total n	4,033

*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 did not target education.
**Wave 1 of the vignette study was attached to the robustness study: disappointment. 595 respondents of the robustness study also participated in the vignette study. The total does not double-count these respondents.

Sample characteristics Table A.1 summarizes the demographic characteristics of each sample.

Exclusion criteria in online experiments Exclusion criteria are preregistered (see Appendix D). The samples do not contain the following responses:

1. Respondents who do not complete the first seven redistribution decisions.[30]
2. Respondents who spend less than 30 seconds on the instructions until the first treatment variation is introduced.
3. Duplicate respondents (very rare cases).

Balanced assignment of experimental conditions Table A.2 and Table A.3 show that the demographic covariates are balanced across experimental conditions in all studies. I test for balanced treatment assignment by regressing the demographic variables on

[30] There is only one redistribution decision in the robustness study. Here, I exclude all respondents who do not complete the study.

a treatment indicator. Across all studies, the coefficient estimates are mostly small, indicating that the demographic covariates are balanced across treatments. For each study, I also test the joint null hypothesis that *all* treatment differences are zero. None of the highly-powered F-test rejects this hypothesis. For the vignette study, the joint effect is marginally significant ($p = 0.083$), but the effect sizes are relatively minor.

Table A.1 Comparison of all samples to the American Community Survey (ACS)

Variable	ACS (2019)	Main study	Equal rates	Disap-pointment	Atten-tion	Attention equal rates	Counter-factual	Vig-nettes
Gender								
Female	51%	51%	52%	63%	52%	48%	53%	61%
Age								
18-34	30%	30%	28%	11%	32%	33%	23%	15%
35-54	32%	33%	32%	30%	32%	29%	35%	33%
55+	38%	37%	41%	59%	36%	38%	42%	52%
Household net income								
Below 50k	37%	40%	43%	47%	39%	44%	39%	45%
50k-100k	31%	34%	32%	34%	34%	33%	32%	33%
Above 100k	31%	27%	26%	19%	26%	23%	30%	22%
Education								
Bachelor's degree or more	31%	43%	40%	48%	38%	36%	56%	47%
Region								
Northeast	17%	21%	16%	25%	16%	16%	17%	25%
Midwest	21%	21%	22%	25%	18%	21%	21%	23%
South	38%	36%	39%	35%	44%	38%	38%	36%
West	24%	22%	23%	15%	23%	25%	24%	16%
Sample size	2,059,945	653	661	606	274	267	945	1,222

Notes: Column 1 presents data from the American Community Survey (ACS) 2019. The other columns describe the different experimental samples.

Table A.2 Test for balanced treatment assignment – part 1

Main study

	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

Robustness study: Equal rates

	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

Robustness study: Disappointment

	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. Each panel represents a study. Within each panel, each column regresses a demographic variable on the 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.3 Test for balanced treatment assignment – part 2

Attention study*

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Midwest	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" study*

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Midwest	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

Counterfactual study

	Female	Age	Income (in $1k)	Bachelor's degree	Region: Midwest	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

*The *Attention* condition of the attention study and the attention "equal rates" study is compared to the *Control* condition of the main study and the robustness "equal rates" study, respectively.

Notes: OLS regressions, robust standard errors in parentheses. Each panel represents a study. Within each panel, each column regresses a demographic variable on the 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 3

Vignette study							
	Female	Age	Income (in $1k)	Bachelor's degree	Region: Midwest	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: Results from the vignette study. OLS regressions, robust standard errors in parentheses. Each column regresses a demographic variable on the treatment dummy to test for imbalanced treatment assignment. A joint F-test, estimated in a SUR model, tests the hypothesis that all treatment differences are zero. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

B Supplementary analyses

B.1 Treatment effects

Average treatment effects in all experimental studies Table B.1 and Table B.2 test for differences in merit judgments across the experimental conditions of the main study, the attention study, the “equal rates” robustness study, the “equal rates” attention study, and the counterfactual study.

Histograms for main and counterfactual study Figure B.1 and Figure B.2 plot the full distribution of reward shares assigned to the disadvantaged worker B in the main study and the counterfactual study, respectively. They show histograms for each experimental condition and each effort scenario.

Heterogeneous treatment effects in main study Table B.3 tests for heterogeneous treatment effects in the main study.

Robustness study: disappointment Table B.4 presents the treatment effects in the robustness study.

Table B.1 Average treatment effects on the reward share of the disadvantaged worker

(A) 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							

(B) Robustness study "Equal rates": Treatment − Control

Effort scenario e	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	2.36	1.06	0.81	-0.16	-0.52	-1.20	0.41	0.39
Standard error	1.38	1.07	0.63	0.18	0.67	1.17	1.25	0.24
CI, 95%	[-0.3, 5.1]	[-1, 3.2]	[-0.4, 2]	[-0.5, 0.2]	[-1.8, 0.8]	[-3.5, 1.1]	[-2, 2.9]	[-0.1, 0.9]
p-values, t-tests	0.088	0.323	0.200	0.364	0.435	0.307	0.745	0.105
p-value, F-test	0.253							

(C) Attention study: Attention − Control

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							

(D) Attention robustness study "Equal rates": Attention − Control

Effort scenario e	0%	10%	30%	50%	70%	90%	100%	Average
Reward diff.	-0.88	0.48	0.20	0.14	-0.21	0.04	0.25	0.00
Standard error	1.23	1.13	0.72	0.21	0.76	1.22	1.33	0.23
CI, 95%	[-3.3, 1.5]	[-1.7, 2.7]	[-1.2, 1.6]	[-0.3, 0.5]	[-1.7, 1.3]	[-2.4, 2.4]	[-2.4, 2.9]	[-0.5, 0.5]
p-values, t-tests	0.473	0.672	0.783	0.509	0.778	0.974	0.850	0.998
p-value, F-test	0.897							

Notes: Results from OLS regressions. Each panel presents the results from a different study. 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. The title of each panel describes which experimental conditions are compared. "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 on the respondent level.

Table B.2 Counterfactual study: Average treatment effects on the reward share of the disadvantaged worker

(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 on the respondent level.

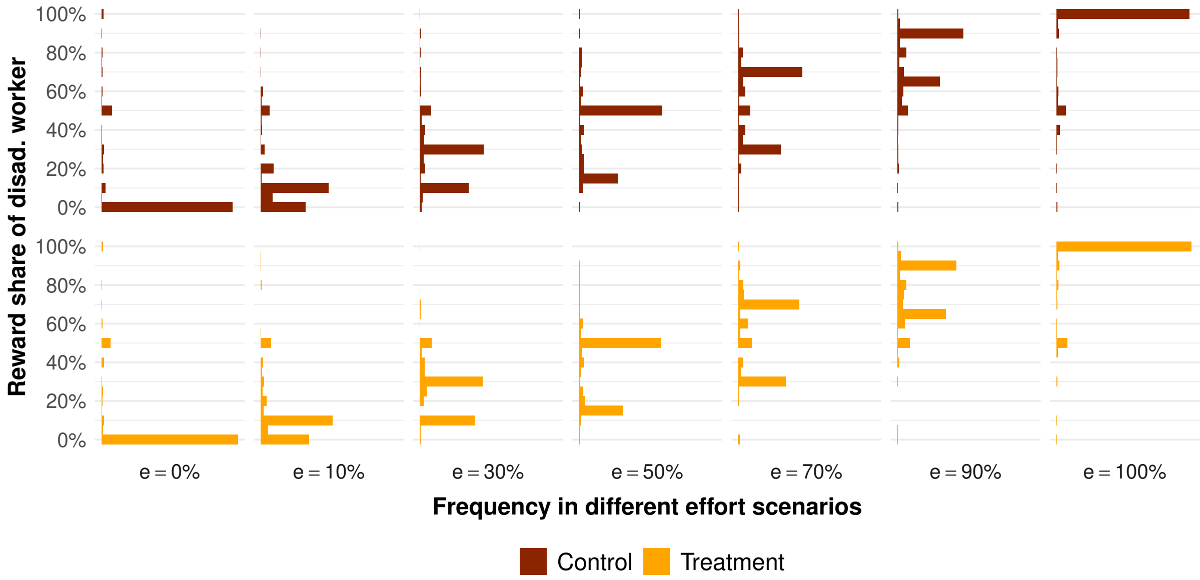


Figure B.1 Main study: Histograms of reward share of disadvantaged worker

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

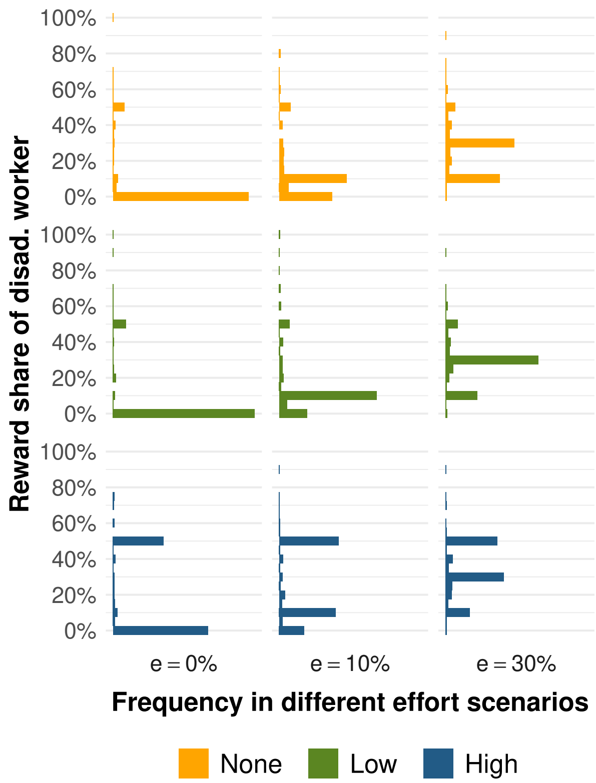


Figure B.2 Counterfactual study: Histograms of reward share of disadv. worker

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 B.3 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.4 Treatment effects in the robustness study: disappointment

	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 robustness study: disappointment, 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.2 Robustness of treatment effects

Robustness of treatment effects Figure B.3 explores the robustness of the treatment effects in the main study, the attention study, the "equal rates" robustness study, the "equal rates" attention study and the counterfactual study. The following robustness specifications are estimated.

1. **Full sample**: Full sample, replicates main results.
2. **Exclude speeders**: I exclude the 25% participants with the lowest response duration.
3. **Exclude quiz mistakes** I exclude participants who answer at least 1 question of the quiz wrongly.
4. **Exclude decisions 4-7** I consider only the first three redistribution decisions of each participant. (Note: Not applicable in the counterfactual study, as I always focus on the first three redistribution decisions here.)
5. **Exclude recognition of real scenario** I drop all respondents who are able to distinguish the hypothetical scenarios from the real one, after they saw all scenarios.
6. **Exclude $0.10** Only applicable to the "equal rates" robustness study and the "equal rates" attention study. The *Control* condition of both studies comes in two variants. Either both workers receive a piece-rate of $0.10 or both respondents receive a piece-rate of $0.50. One concern is that only the latter variant can be cleanly compared to the *Treatment* condition in which both workers end up with a piece-rate of $0.50. This robustness check therefore excludes spectators in the *Control* condition with a piece-rate of $0.10.

The estimated treatment effects are robust in all studies.

Robustness to the order of workers In the experiment, I randomize whether worker A or worker B is advantaged or disadvantaged. The main analysis recodes all responses as if A was the advantaged worker to ease analysis and exposition. Here, I test whether a reverse order of workers, that is a worker pair in which worker A is disadvantaged and worker B is advantaged, affects merit judgments. I regress the average reward share respondents assign to the disadvantaged worker on a dummy for reversely ordered worker pairs. Table B.5 shows the results. The random variation in the order of workers does not affect merit judgments.

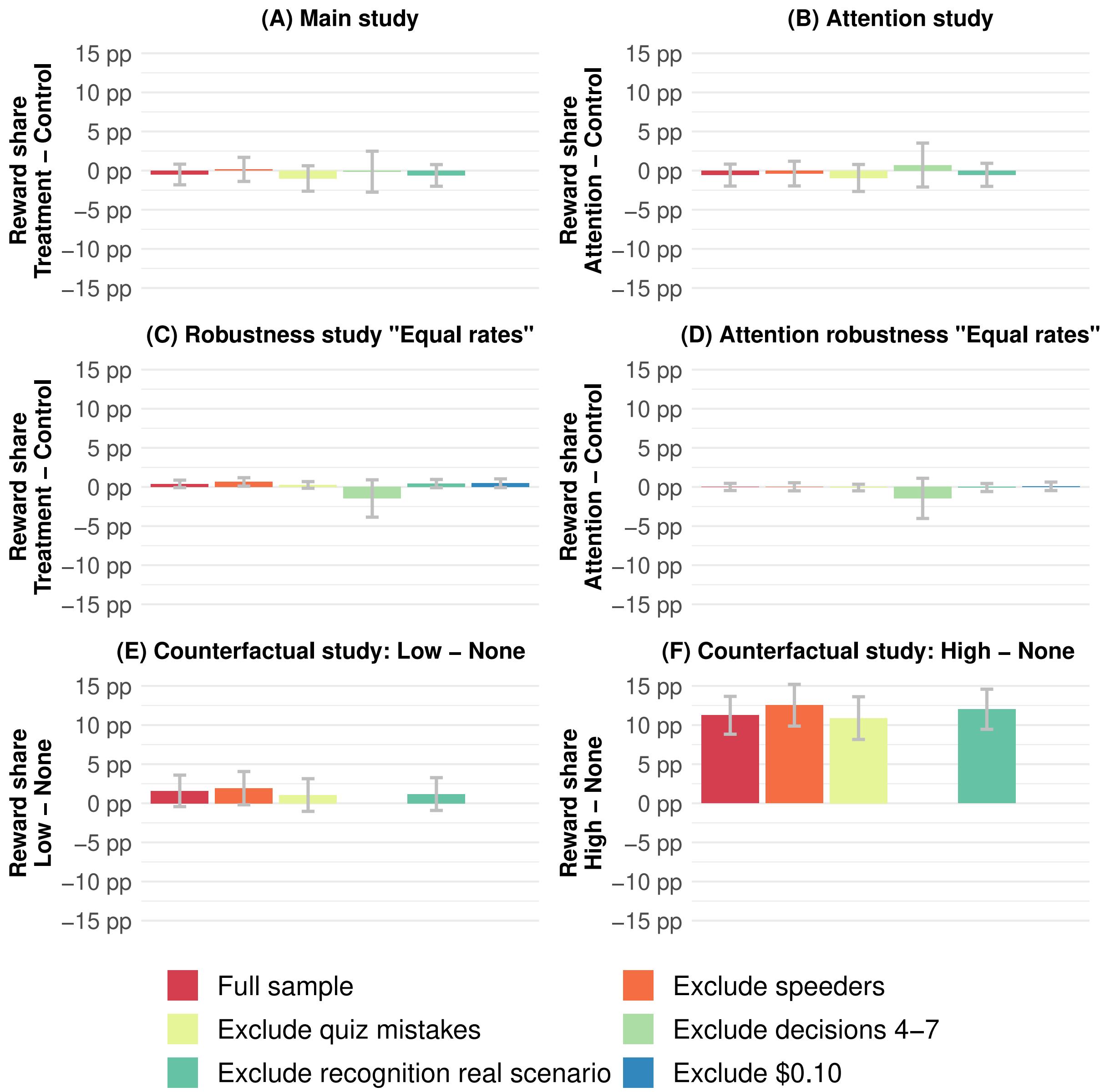


Figure B.3 Robustness of average treatment effects (with 95% CI)

Notes: Results from the main, attention, "equal rates" robustness, "equal rates" attention, and counterfactual studies. Each panel presents the results from a different study. Each panel plots the treatment effect on the reward share assigned to the disadvantaged worker B (averaged across the effort scenarios) in different robustness specifications. See above for a description. The gray errorbars are 95% confidence intervals.

Table B.5 Robustness of merit judgments to the order of workers

	Mean reward share of disadv. worker (in %)				
	Main study	Robustness study "Equal rates"	Attention study	Attention study "Equal rates"	All
	(1)	(2)	(3)	(4)	(5)
Reverse order	−0.327	0.133	−0.058	0.274	−0.037
	(0.674)	(0.243)	(1.064)	(0.344)	(0.302)
Condition FE	✓	✓	✓	✓	✓
Observations	653	661	274	267	1,855
R^2	0.001	0.004	0.000	0.002	0.186

Notes: Results of the main study, the "equal rates" robustness study, the attention study, and the "equal rates" attention study. OLS regressions, robust standard errors in parentheses. The outcome variable is the reward share assigned to the disadvantaged worker, averaged across all seven effort scenarios. The independent variable is a dummy that takes value 1 if worker A is disadvantaged and worker B is advantaged and value 0 for the opposite case. (Note: In the remainder of the paper, I recode all responses as if A was the advantaged worker to ease analysis and exposition.) Columns 1-4 present results from different studies. Column 5 presents a pooled estimate. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.3 Beliefs about situational influence in the main study

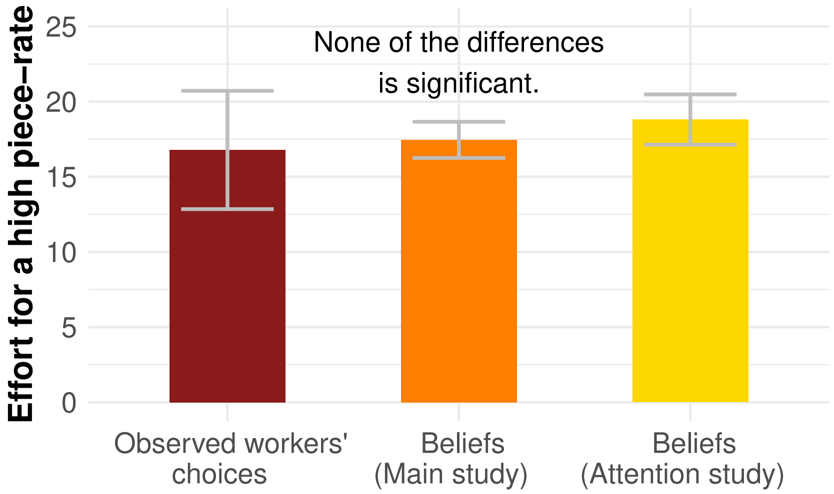


Figure B.4 Average beliefs about the piece-rate effect (with 95% CI)

Notes: Results from the main and the attention study. 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 study. The gray errorbars are 95% confidence intervals. t-tests are used to evaluate the significance of the differences.

B.4 Structural model of merit views

Maximum likelihood estimation

Data Counterfactual study, decisions 4-7, 945 respondents. In decisions 4-7, respondents face a randomly generated effort scenario.[31] 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 $\{0, 1, ..., 49, 50\}$.
- Effort of worker B: Uniformly randomly drawn from the set $\{0, 1, ..., 24, 25\}$.
- Counterfactual effort of worker B for a high piece-rate: The difference between the counterfactual and observed effort is uniformly randomly drawn from the set $\{0, 1, ..., 24, 25\}$.
- The effort and initial payment shares of both workers follow from the above variables.

In the baseline condition, no information about the counterfactual effort choice of the disadvantaged worker is provided. In the "low counterfactual" and the "high counterfactual" conditions, spectators are informed about what the disadvantaged worker would have made in advantaged circumstances.

Model Each individual endorses one of the four merit views that are discussed in Section 2 of the main text. A respondent i of type t rewards the workers according to her merit view $m_{t(i)}(e, s)$ in scenario s and a normally distributed response error $\varepsilon_{is} \sim_{iid} N(0, \sigma^2)$. That is, $r_{is} = m_{t(i)}(e, s) + \varepsilon_{is}$.

As discussed in Section 6, I parametrize the fairness view of comparable choice meritocrats (CCM) as follows:

$$m_{\text{CCM}}(e, s) = \begin{cases} \rho Ec(e, s) + (1 - \rho)e & \text{if counterfactual is uncertain} \\ c(e, s) & \text{if counterfactual is known} \end{cases}$$

This means that comparable choice meritocrats tend to discount the expected counterfactual effort choice if it is uncertain. The discount parameter is ρ.

I also need to estimate spectators' expectation of the counterfactual effort share, $Ec_(e, s)$, when the counterfactual is unknown. In line with the evidence of Section 5, I

[31] The contingent response method allows me to freely vary the effort choices of workers in the hypothetical scenarios without being deceptive.

assume that spectators correctly anticipate the average effect of the piece-rate. Moreover, I assume that the real piece-rate effect is constant in line with the discussion in Section 2. In the data, I observe that workers are willing to complete about 12.5 tasks more for a high piece-rate (see Table C.1, Column 3). I use this estimate to derive spectator's expected counterfactual effort choice of worker B ($EC_B = E_B + 12.5$) for each effort scenario in the baseline condition where the counterfactual effort choice is unknown.[32] This allows me to derive $Ec(e,s) \approx \frac{EC_B}{E_A+EC_B}$. Below, I show that I obtain virtually identical result with an alternative specification of $Ec(e,s)$. The results are insensitive to the calibration $Ec(e,s)$ because the spectators fully discount it anyway.

I estimate six parameters: the population shares $\boldsymbol{\theta}$ of the four merit views ($\sum_t \theta_t = 1$), the discount parameter ρ, and the standard deviation of the response error σ. I impose $0 \leq \theta_t \leq 1 \; \forall t$, $0 \leq \rho \leq 1$, and $\sigma > 0$.

Log-likelihood

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

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

$$\Pr(\boldsymbol{r}_i \mid \theta_t, \rho, \sigma) = \prod_s \varphi(r_{is} - m_{t(i)}(s, e, \rho), \sigma^2) \tag{3}$$

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 ρ, σ, and the share of actual choice meritocrats, comparable choice meritocrats, and libertarians. The share of egalitarians follows via $\sum_t \theta_t = 1$.

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.

Inference for constrained maximum likelihood Standard inference in constrained maximum likelihood models can become unreliable if one of the parameters is on or

[32]Workers can complete at most 50 tasks, so I cap the counterfactual effort choices at 50.

near the boundary (Schoenberg, 1997). Since I estimate a ρ of 0.00 which is on the boundary, caution seems to be warranted. The discussion below indicates, however, that the inference is nevertheless reliable.

First, I obtain virtually identical estimates and standard errors for $\boldsymbol{\theta}$ and σ if I estimate the model without constraints (results available upon request).

Moreover, I assess the coverage of the confidence intervals in an independent simulation experiment. To this end, I generate 1,000 simulated data sets from the model, assuming that the main estimates in Table 5 are the true parameter values. In particular, I impose $\rho = 0$. For each simulated data set, I derive the maximum likelihood estimates and their associated 95% confidence intervals. Then, I assess whether the confidence intervals cover the "true" parameters in about 95% of cases. This is indeed the case. The estimated coverage frequency ranges from 93.4% to 97.2%. I obtain similar results if I randomly perturb $\boldsymbol{\theta}$ and σ in each simulation to explore the coverage in the neighborhood of the estimated parameters (here, the coverage ranges from 94.5% to 98.8%).

Robustness of estimates

Table B.6 shows that the results of the maximum likelihood are robust across several different specifications.

- Main: Main specification
- 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.
- Multipl. effort: Here, I calibrate spectators' expectations of worker B's counterfactual effort as $EC_B = 3.3 * E_B$, assuming that the effect of the higher piece-rate is multiplicative. In the data, I observe that workers are willing to complete about 3.3 as many tasks for a high piece-rate than for a low piece-rate (see Table C.1, Column 3).
- Bounds adjust: Because the support of normal noise is unbounded, the likelihood function assigns 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%].

- Trembling: I explore an alternative error specification. 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

The model allows to estimate whether its parameters differ for subgroups of respondents. Consider two groups of respondents A and B. I assume that the population shares of different fairness types and the counterfactual discount parameter are $(\boldsymbol{\theta},\ \rho)$ in group A. In group B, the population shares are $(\boldsymbol{\theta},\ \rho)+\boldsymbol{\lambda}$. 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 B.7 displays the resulting estimates of $\boldsymbol{\lambda}$.

Table B.6 Robustness of structural estimation

	(1) **Main**	(2) **Duration**	(3) **Quiz**	(4) **Guess correct**	(5) **Multipl. effort**	(6) **Bounds adjust**	(7) **Trembling**
Population shares							
Actual choice meritocrats	36.7%	35.5%	39.9%	36.8%	36.7%	35.9%	34.7%
	(1.9%)	(2.1%)	(2.3%)	(2.0%)	(1.9%)	(2.1%)	(1.9%)
Comparable choice meritocrats	26.2%	28.9%	27.3%	26.2%	26.2%	26.2%	29.2%
	(1.7%)	(2.0%)	(2.1%)	(1.9%)	(1.8%)	(2.1%)	(1.8%)
Libertarians	23.0%	23.7%	22.5%	23.4%	23.0%	23.8%	24.8%
	(1.4%)	(1.6%)	(1.7%)	(1.5%)	(1.4%)	(1.4%)	(1.5%)
Egalitarians	14.2%	11.9%	10.4%	13.6%	14.1%	14.1%	11.3%
	(–)	(–)	(–)	(–)	(–)	(–)	(–)
Counterfactual discount parameter							
ρ	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.04)	(0.05)	(0.05)	(0.04)	(0.06)	(0.11)	(0.01)
Error term and sample							
σ noise	9.27	9.16	8.60	9.32	9.27	9.72	
	(0.11)	(0.13)	(0.12)	(0.12)	(0.11)	(0.13)	
α noise							0.23
							(0.01)
Respondents	945	708	656	834	945	945	945
Decisions	3777	2831	2621	3333	3777	3777	3777

Notes: Results from counterfactual study, decisions 4-7. Maximum likelihood estimation of the structural model of merit views. Standard errors in parentheses. The estimates indicate the population shares of different fairness views and the discounting parameter ρ. 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.

Table B.7 Differences of 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	1.7%	-2.2%	0.7%	7.0%*
	(3.7%)	(3.8%)	(3.8%)	(3.8%)
Comparable choice meritocrats	1.6%	0.8%	-1.5%	-0.2%
	(3.5%)	(3.5%)	(3.5%)	(3.6%)
Libertarians	-1.3%	2.4%	1.9%	-4.1%
	(2.9%)	(2.9%)	(2.9%)	(2.9%)
Egalitarians	-2.0%	-1.0%	-1.1%	-2.8%
	(–)	(–)	(–)	(–)
Differences in counterfactual reasoning				
ρ	0.00	0.00	0.00	0.00
	(0.09)	(0.10)	(0.09)	(0.09)
Sample				
Respondents	916	916	916	916
Decisions	3661	3661	3661	3661

Notes: Results from counterfactual study, decisions 4-7. Maximum likelihood estimation of the structural model of merit 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 ($\boldsymbol{\theta}$ and ρ) 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. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

B.5 Vignette study

Robustness of treatment effects Table B.8 shows that the results of the vignette study are largely insensitive to the exclusion criterion and to 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).

- **Main**: Main specification
- **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.

Results of additional crime vignette The vignette survey also contained a fourth vignette on criminal behavior (see Appendix F for the full vignette wording).

> **Crime vignette**: In this vignette, the advantaged person grew up in a rich neighborhood with low crime rates. He went to good schools, and his parents made sure he grew up in a loving, nurturing environment. The disadvantaged person grew up in a poor neighborhood with very high crime rates. His parents often neglected him, and both his family and peers committed crimes. While the advantaged person started studying business and works as a salesman, the disadvantaged person started selling drugs and frequently violates the law. Both earn $50,000 a year today.

In contrast to the other vignette, the crime vignette revolves around legal versus illegal behavior instead of hard work or entrepreneurial risk-taking, and both persons earn equal instead of unequal incomes. As a consequence, respondents redistribute money *away* from the disadvantaged, criminal person in the baseline condition, likely because they reject the illegal source of his income. Only 41% accept the initial income equality between both persons (Column 1, Table B.9). This fraction is virtually identical in the low counterfactual treatment, but 12.3 percentage points higher in the high counterfactual treatment, replicating the findings in the other vignettes.

Still, Column 2 suggests that the average reward share of the unlawful person might be slightly lower when respondents know that the person would violate the law even if he had grown up in privileged circumstances. This effect is driven by a slightly larger share of respondents who take all money away from the unlawful person (Column 3). Both effects are however only marginally significant.

Table B.8 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 on 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 that are explained above. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

Table B.9 Vignette study: Results from the crime vignette

	Binary indicator for equal shares	Reward share of disadv. person (in %)	Binary indicator for giving 0% to the disadv. person
	(1)	(2)	(3)
Low counterfactual	-0.031	-3.066^{*}	0.056^{*}
	(0.040)	(1.649)	(0.029)
High counterfactual	0.123^{***}	3.347^{**}	-0.004
	(0.040)	(1.571)	(0.027)
Constant	0.412^{***}	34.111^{***}	0.124^{***}
	(0.028)	(1.114)	(0.019)
Observations	894	894	894
R^2	0.018	0.017	0.006

Notes: Results from the vignette study. OLS regressions with robust standard errors. Column 1 regresses a binary indicator for whether a respondent accepts the reward equality between both persons on treatment dummies. The dependent variable in Column 2 is the reward share assigned the disadvantaged person. In Column 3, the dependent variable is a binary indicator for taking all money away from the unlawful person. *** $p < 01$, ** $p < 0.05$, * $p < 0.10$.

C Endogenous effort choices 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 548 workers who were recruited for the study. 336 workers were recruited for the *main*, *robustness "equal rates"*, *attention*, and *attention "equal rates"* studies (Amazon Mechanical Turk, US, May and June 2020). 212 were recruited for the *counterfactual* study (Amazon Mechanical Turk, US, January 2021).[33]

Table C.1 regresses the number of completed tasks 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, Robustness "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.

The higher piece-rate leads to a 333% higher effort in the main condition, a 155% higher effort in the robustness "equal rates" condition, and a 335% higher effort in the counterfactual condition. Thus, the external piece-rate strongly affects how much effort the workers exert.

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

	Effort (number of completed tasks)		
	Main	Robustness "Equal rates"	Counterfactual
	(1)	(2)	(3)
High rate	11.744***	5.553**	12.547***
	(2.308)	(2.357)	(1.540)
Constant	5.040***	10.044***	5.349***
	(1.135)	(1.226)	(1.043)
Observations	124	212	212
R^2	0.142	0.029	0.149

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

[33]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.

D Research transparency

Preregistration The main study, the robustness study: equal rates, the robustness study: disappointment, the attention study, the attention "equal rates" study, and the 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 uses a different title and different treatment labels.
- Non-preregistered analyses include the comparison of worker B's reward share, averaged across effort scenarios (a straight-forward 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 size differs slightly from the pre-registered size of about 300 per condition due to the logistics of the sampling process.
- 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 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 vignette study was not pre-registered.

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.

E Extract from the main study's instructions

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

Part 1

In what follows, we will ask you to make a series of decisions that might have **consequences for a real-life situation**.

Please read the following pages very carefully. A **quiz** will test your understanding. You can proceed with the study only if you answer all quiz questions correctly.

→

– PAGE BREAK –

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.

← →

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.

← →

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.

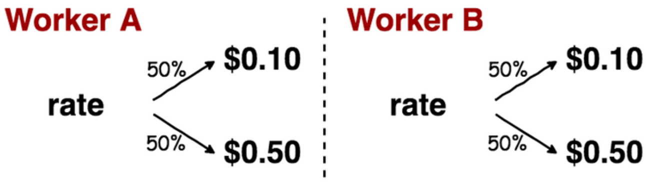
← →

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.



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.

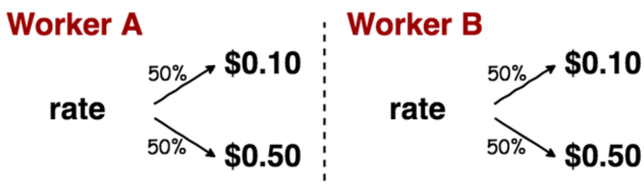
← →

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.



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 –

Scenario 1

	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 –

Scenario 1

	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	%

F Extract from the vignette study's instructions

This appendix shows the scenario descriptions from the vignette study. The full instructions for the vignette study are available at https://osf.io/xj7vc/.

F.1 Scenario "discrimination"

Richard and Oliver work for the same company. In the last months, they competed for a promotion that came with an attractive one-time bonus of $10,000.

However, their boss is notorious for favoring white employees. In fact, he has never promoted a black person before, although he has had plenty of opportunities to do so.

Richard is white. He worked hard to win the promotion.

Oliver is black. He did not work hard to win the promotion.

Who got promoted?

As a consequence of their choices, Richard is promoted and receives the bonus of $10,000. Oliver is not promoted and receives no bonus.

[*Addendum, High counterfactual condition*]

What if the boss did not favor white employees?

Assume that if the boss did not favor white employees, Oliver would have made the same choice as Richard. **Oliver would have worked as hard as Richard did.**

[*Addendum, Low counterfactual condition*]

What if the boss did not favor white employees?

Assume that if his boss did not favor white employees, Oliver would still have made the same choice. **Oliver would not have worked hard.**

F.2 Scenario "poverty"

Mike

Mike grew up in a rich family. He was always told, "In this country, you can go as far as your hard work takes you." His family expected him to work hard. Mike went to good, engaging schools that challenged him. He knew he would be popular among his peers if he achieved good grades and worked hard.

Mike has always worked hard in his life.

Paul

Paul grew up in a poor family. He was always told, "In this country, the poor stay poor, and the rich get richer." His family did not expect him to work hard. Paul went to poor-quality schools where he was bored and never challenged. He knew he would be popular among his peers if he was lazy, rebelled against authority, and violated rules.

Paul has never worked hard in his life.

Income today

As a consequence of their choices, Mike earns $125,000 a year, and Paul earns $25,000 a year.

[*Addendum, High counterfactual condition*]

What if Paul had grown up in Mike's environment?

Assume that if Paul had grown up in the same environment as Mike, he would have made the same choices as Mike. **Paul would always have worked as hard as Mike did.**

[*Addendum, Low counterfactual condition*]

What if Paul had grown up in Mike's environment?

Assume that if Paul had grown up in the same environment as Mike, he would still have made the same choices. **Paul would never have worked hard in his life.**

F.3 Scenario "start-up"

Frank

Frank always dreamed of founding his own software start-up. He knew that he would **inherit a considerable fortune**. Therefore, he knew that he had enough money to launch his start-up, and that even if his first attempts failed, he would have enough money left to try again and pursue a new business idea.

Frank decided to take the risk and founded his own software start-up.

Ray

Ray always dreamed of founding his own software start-up, too. However, Ray's parents were poor and he had **very little money**. Therefore, he knew that it would be difficult to find enough money to launch a start-up, and he knew that if his first attempt failed, he would be broke.

Ray decided not to take the risk. Instead, he works as a software developer for a local company.

Income today

As a consequence of their choices, Frank earns $200,000 a year, and Ray earns $50,000 a year.

[*Addendum, High counterfactual condition*]

What if Ray had had as much money as Frank?

Assume that if Ray had had as much money as Frank, he would have made the same choices as Frank. **Ray would have taken the risk and founded his own software start-up.**

[*Addendum, Low counterfactual condition*]

What if Ray had had as much money as Frank?

Assume that if Ray had had as much money as Frank, he would still have made the same choices. **Ray would have decided not to take the risk. Instead, he would work as a software developer for a local company.**

F.4 Scenario “crime”

Robert

Robert grew up in a rich neighborhood with very low crime rates. His parents made sure he grew up in a loving, nurturing environment. Robert has always been told, “In this country, you can rise as far as you want if you play by the rules.” Robert went to good, engaging schools that challenged him. Many of his peers planned to study at a university.

Robert started studying business at the age of 20. Today, he works as salesman. He never does anything illegal.

John

John grew up in a poor neighborhood with very high crime rates. His parents often neglected him. Once his father was caught selling drugs and had to spend several years in jail. John has always been told, “Playing by the rules means nothing when the rules are stacked against you.” He went to poor-quality schools where he was bored and never challenged. Many of his peers had already committed crimes by the time they reached their teenage years.

John committed his first crime at the age of 20. Today, he sells drugs. He frequently violates the law.

Income today

As a consequence of their choices, Robert earns $50,000 a year, and John earns $50,000 a year.

[*Addendum, High counterfactual condition*]

What if John had grown up in Robert’s environment?

Assume that if John had grown up in the same environment as Robert, he would have made the same choices as Robert. **John would never do anything illegal.**

[*Addendum, Low counterfactual condition*]

What if John had grown up in Robert’s environment?

Assume that if John had grown up in the same environment as Robert, he would still have made the same choices. **John would sell drugs and frequently violate the law.**