Experimental Evidence on the Acceptance of Males Falling Behind

BY Alexander W. Cappelen, Ranveig Falch and Bertil Tungodden

DISCUSSION PAPER

Institutt for samfunnsøkonomi
Department of Economics

SAM 13/2023

ISSN: 0804-6824

June 2023

Experimental Evidence on the Acceptance of Males Falling Behind

ALEXANDER W. CAPPELEN
Norwegian School of Economics
RANVEIG FALCH
Norwegian School of Economics
BERTIL TUNGODDEN
Norwegian School of Economics

Abstract

In recent decades, an increasing share of males struggle in the labor market and education. We show in a set of large-scale experimental studies involving more than 30,000 Americans that people are more accepting of males falling behind than females falling behind and less supportive of government policies supporting males falling behind. We provide evidence of the underlying mechanism being statistical fairness discrimination: people consider males falling behind to be less deserving of support than females falling behind because they believe that males are more likely than females to fall behind due to lack of effort. The findings are important for understanding how society perceive and respond to the growing number of disadvantaged males.

Key words: inequality, statistical fairness discrimination, experiment

JEL Codes: C91, D63, J16

*Corresponding author; E-mail: bertil.tungodden@nhh.no. Phone: +47 55959261, Department of Economics, NHH Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway.

1 Introduction

Males continue to dominate top-level positions across the world, and there is still a significant gender wage gap in all societies. In the United States, female full-time workers earned on average 83 percent of what men did on an annual basis in 2021 (U.S. Bureau of Labor Statistics, 2022). Consequently, there is an urgent need for political action to achieve gender equality for women.

The present paper is motivated by a different gender gap that has emerged in the labor market and education in recent decades (Autor and Wasserman, 2013). In high-income countries, there is growing concern about the prospects for low-skilled males: "The decline in economic opportunities for low-skilled men and the possible negative effects of this trend on their well-being is a matter of increasingly urgent concern for policy makers and the general public" (Coile and Duggan, 2019, p. 2). Males with less than a four-year college education have experienced a significant reduction in real income over the last decade in the US (Autor and Wasserman, 2013; Binder and Bound, 2019), and the percentage of young males and prime-age males outside the labor force has increased (Blau and Kahn, 2013; Krueger, 2017). The prospects for males outside the labor force are dim, in particular for those from low-income households and for males with minority backgrounds. The likelihood of living in poverty has increased and their expected future health and emotional well-being is poor (Autor and Wasserman, 2013; Council of Economic Advisers, 2016; Krueger, 2017).

There is a also growing concern about boys falling behind in education. In most OECD countries, a larger share of boys than girls do not attain the baseline level of proficiency in any of the core subjects; mathematics, reading and science (OECD, 2015). In the US for instance, the average percentage of students who do not attain the baseline proficiency level was 71% higher for boys than for girls. Boys are also dropping out of high school at higher rates than girls in most OECD countries. In higher education, females have surpassed the rate of males graduating in nearly all OECD countries, on average by 12 percentage points (OECD, 2022).

These striking developments make it important to study how people perceive and react to males falling behind, which is likely to shape policies targeting disadvantaged males both in the public and private sector. In this paper, we report from a large-scale experimental study of how people's fairness preferences and beliefs depend on the gender of the person falling behind. In particular, we are interested in whether there is statistical fairness discrimination against males falling behind: do people consider males falling behind to be less deserving of support than females falling behind because they believe that males are more likely than females to fall behind due to lack of effort? In the first part of the paper, we provide choice experimental evidence on whether people are more accepting of inequalities when males fall behind than when females fall behind in a controlled work environment, and whether people to a greater extent believe that males fall behind due to lack of

effort compared to females in this environment. In the second part, we report from a corresponding policy experiment to study whether the findings in the choice experiment carry over to the policy domain. In total, more than 30,000 Americans participated in the different experiments in this study.

To gather choice experimental data from a representative sample of the general population in the United States, we leverage the infrastructure of a leading international data-collection agency along with an online labor market (Almås, Cappelen, and Tungodden, 2020). In the online labor market, we recruit workers and create inequalities by paying two workers differently for the same assignment. In our main treatments, the inequality is generated by paying the more productive worker more than the less productive worker. We then ask a general population sample of Americans to act as third-party spectators and make consequential decisions about whether to redistribute earnings between the two workers. Spectators are randomly assigned to a treatment where the low productive worker is a male or to a treatment where the low productive worker is a female. Our main interest is in studying whether the redistribution decisions of the spectators depend on the gender of the low productive worker. We run a set of additional treatments to shed light on the underlying mechanisms of the spectators' choices, where we vary the source of the inequality (merit or luck) and the gender composition of the two workers (mixed-gender or single-gender), and, in an independent general population sample of Americans, elicit beliefs about the effort exerted by the males and females who fall behind in the choice experiment. To study statistical fairness discrimination in the policy domain, we implement a large-scale survey experiment that investigates whether support for government policies targeting people falling behind depends on the gender of those falling behind, and whether beliefs about effort are different for males and females falling behind in the labor market and education.

Our findings provide compelling evidence of males falling behind being treated differently from females falling behind, both in the choice experiment and in the policy experiment, and of the underlying mechanism being statistical fairness discrimination. The main findings are illustrated in Figure 1. In a controlled work environment, a significantly higher share of people choose not to transfer anything to a male falling behind due to low productivity than to a female falling behind, 38.4% versus 31.1% (Panel A), and a significantly higher share of people believe that males fall behind in the choice experiment due to lack of effort than females, 53.0% versus 44.3% (Panel B). These findings are similarly reflected in the policy experiment: fewer people express support for government assistance to males falling behind than to females falling behind in the labor market and education, 54.2% versus 42.2% (Panel C), and more people believe that males fall behind in the labor market and education due to lack of effort than females, 46.7% versus 32.5% (Panel D). The findings are robust across subgroups in society, to different experimental manipulations and empirical specifications, and to multiple hypothesis testing. We also find that the treatment effect of

manipulating the gender in the choice experiment is comparable to the treatment effect of manipulating the gender in the policy experiment. The magnitude of the gender treatment effect is in most cases comparable to the difference between Republicans and non-Republicans. Taken together, the evidence strongly suggests that statistical fairness discrimination is an important mechanism for understanding how Americans relate to males falling behind.

[Figure 1 about here]

Our paper contributes to several important strands of literatures. The findings augments the growing literature on the negative developments of males in the labor market and education. The existing literature has documented how the low performance of males reflects structural changes and socioeconomic developments (Almås, Cappelen, Salvanes, Sørensen, and Tungodden, 2016; Autor and Wasserman, 2013; Autor, Figlio, Karbownik, Roth, and Wasserman, 2019; Bertrand and Pan, 2013; Binder and Bound, 2019; Krueger, 2017; Rosin, 2012), and the present paper complements these findings by providing novel evidence on how society perceives males who fall behind. We show that people believe that it is more likely that males falling behind have exerted low effort than females falling behind, which may result in males falling behind being treated differently in school, at the workplace, and in the family.

We further contribute to the literature on gender discrimination (Bertrand and Duflo, 2017), by showing how males who fall behind may face statistical fairness discrimination. While important papers have documented discrimination against females in different domains, such as in hiring decisions (Bohren, Haggag, Imas, and Pope, forthcoming; Goldin and Rouse, 2000; Coffman, Exley, and Niederle, 2021), task allocation (Babcock, Recalde, Vesterlund, and Weingart, 2017), bargaining (Castillo, Petrie, Torero, and Vesterlund, 2013; Exley, Niederle, and Vesterlund, 2020), teaching evaluations (Mengel, Sauermann, and Zölitz, 2018), and career development (Reuben, Sapienza, and Zingales, 2014), recent studies have also highlighted discrimination against males in certain settings (Bohren, Imas, and Rosenberg, 2019; Mengel et al., 2018; Reynolds, Howard, Sjåstad, Zhu, Okimoto, Baumeister, Aquino, and Kim, 2020; Williams and Ceci, 2015). Notably, Bohren et al. (2019) find in a field experiment on an online mathematics forum for STEM-students and researchers that females initially encounter significant discrimination, but gradually become favored over men. Our paper focuses specifically on males who fall behind in a distributive context. In a large-scale study of the general population in the US, we demonstrate that people are less willing to support males falling behind than females falling behind. Furthermore, we contribute to the discrimination literature by presenting evidence of the underlying mechanism being statistical fairness discrimination, with people in their fairness considerations making inferences about whether a worker is deserving of support based on their background characteristics. Statistical

fairness discrimination is likely to extend beyond the distributive setting explored in our study, and may be relevant for understanding discriminatory behavior in any context where a principal would like to reward effort.

Finally, our results contribute to the literature in behavioral economics, specifically in understanding the role of gender considerations in people's social preferences (Croson and Gneezy, 2009; Eckel and Grossman, 2008). Prior studies have investigated the effect of varying the salience of the recipient's gender in dictator games and ultimatum games. A meta-study on dictator game experiments finds that females receive higher allocations compared to males in dictator games when the gender of the recipient is made salient (Engel, 2011). In ultimatum games that manipulate the gender of the recipient, individuals tend to make higher offers to males compared to females (Eckel and Grossman, 2001; Solnick, 2001). In contrast to these earlier studies, we examine gender discrimination in third-party spectator choices, which provides a direct expression of participants' moral preferences (Cappelen, Konow, Sørensen, and Tungodden, 2013). In this setting, we demonstrate significant statistical fairness discrimination against males. We also provide new evidence on how fairness preferences shape distributive behavior (Bolton and Ockenfels, 2000; Bortolotti, Soraperra, Sutter, and Zoller, 2017; Cappelen, Drange Hole, Sørensen, and Tungodden, 2007; Cappelen, Falch, and Tungodden, 2020; Falch, 2022; Fehr and Schmidt, 1999; Hvidberg, Kreiner, and Stantcheva, forthcoming; Konow, 2000), by showing how people's fairness considerations differ across contexts and the characteristics of the workers. Finally, the study contributes to the behavioral literature by providing unique large-scale evidence on how gender influences inequality acceptance in a general population sample.

The paper is organized as follows: Section 2 presents the experimental designs, Section 3 outlines the empirical strategy. The main results from the choice experiment are reported in Section 4, while the findings from the policy experiment are presented in Section 5. Section 6 concludes. Additional analysis and the complete instructions are provided in the online appendix.

2 Research design and participants

We report from two large-scale experimental studies examining people's acceptance of males falling behind. The first study is a *choice experiment* where third-party spectators decide whether to redistribute from a more productive worker with high earnings to a less productive worker with no earnings. The second study is a *policy experiment* where participants state their level of agreement with the government providing support to people falling behind in the labor market and education. Both studies employ a between-subject design where we manipulate the gender of those falling behind, which allows us to causally identify whether people are less concerned about males falling behind compared to females falling behind. To study whether statistical fairness discrimination

can explain the observed patterns, we also elicit beliefs about the effort exerted by those who fall behind.[1]

2.1 Choice experiment

Table 1 summarizes the main stages of the spectator design in the choice experiment, which builds on Almås et al. (2020). First, *workers* were recruited to complete an assignment.[2] Second, the workers were matched in pairs and assigned different earnings. Third, the *spectators* were randomly matched to a pair of workers and decided whether to redistribute earnings between the two workers. Finally, the workers were paid according to the spectators' decisions.

[Table 1 about here]

In all treatments, the spectators were presented with a situation where one worker had earnings of 6 USD and the other had earnings of 0 USD. They were informed that both workers were from the US and of the same age. The spectators were not informed about the nature of the tasks assigned to the workers. It was emphasized to the spectators that, unlike traditional survey questions, their choice was consequential. To minimize the role of worker expectations, the spectators were told that the workers would not at any point be informed about their initial earnings.

In the two main treatments, referred to as the mixed-gender merit treatments, the spectators consider a distributive situation involving a female worker and a male worker. The initial inequality in earnings in the merit treatments is determined by the productivity of the workers: the more productive worker earns 6 USD and the less productive worker earns 0 USD. The only difference between the two merit treatments is whether the low productive worker is a male or a female. This experimental design allows us to causally identify whether spectators are more inequality accepting when a male is falling behind compared to when a female is falling behind. To provide further evidence on how the source of inequality affects the spectators' concern for the worker falling behind, we also implemented two mixed-gender luck treatments that mirrored the two main treatments, but where the inequality in earnings was determined by luck. Finally, to investigate the

[1] This study was implemented through a series of data collections, in collaboration with international survey providers. See Online Appendix B.1 for further details. We implemented three rounds of the choice experiment, with a pre-analysis plan specified for each round (registered at the AEA RCT Registry: AEARCTR-0000853, AEARCTR-0001027 and AEARCTR-0005610). The main analysis is in line with the pre-registered empirical strategy and hypotheses. In Online Appendix A.3, we provide a discussion of how and when we deviate from the pre-analysis plans. We did not specify separate pre-analysis plans for the other parts of the study, since they mirror the spectator design in the choice experiment. The instructions for the different parts of the study can be found in Online Appendix B.

[2] The workers were recruited from the online labor market platform Amazon Mechanical Turk, which specializes in recruiting anonymous workers to complete small tasks online. Upon recruitment, the workers were guaranteed a participation fee of 2 USD and informed that they could earn additional money based on their own and others' decisions. The sole function of the workers is to render the spectators' decisions consequential.

effect of the gender composition of the worker pair on the spectators' behavior, we implemented four single-gender treatments that mirrored the four mixed-gender treatments. We employed a between-subject design, where spectators were randomly assigned to treatments, see Table S.1 for an overview.

To examine beliefs about the effort exerted by the worker falling behind, we conducted an independent survey involving a distinct sample of participants who were randomly assigned to one of the treatments in the choice experiment.[3] The participants were asked to state the extent to which they agreed with the worker with no earnings, in expectation, having exerted less effort than the worker with earnings, the response scale being from 1 (strongly disagree) to 5 (strongly agree). This survey design allows us to causally identify, within the controlled work environment of the choice experiment, whether people believe that males are more likely than females to fall behind due to lack of effort.

2.2 Policy experiment

In the policy experiment, participants were asked whether they agreed with the government providing support to people falling behind in the labor market and education. In a between-subject design, we randomly manipulated the gender of those who fall behind. Specifically, the participants were presented with the following statement: "It is very important that the government provides support to males (females) who fall behind in the labor market and education," and then responded on a scale from 1 (strongly disagree) to 5 (strongly agree).

To elicit beliefs about the role of effort of people falling behind in the labor market and education, we employed two different approaches. In one study, we asked participants about the role of effort in *explaining* why some fall behind: "We observe some males (females) falling behind in education and in the labor market. To what extent do you agree with the statement: When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort." In a second study, we asked directly about their beliefs about the *level of effort* of people falling behind: "We observe some males (females) falling behind in education and in the labor market. To what extent do you agree with the statement: Males (Females) falling behind in education and in the labor market have exerted low effort." For both questions, the participants responded on a scale from 1 (strongly disagree) to 5 (strongly agree). The two questions allow us to identify whether people are more likely to believe that males fall behind due to lack of effort in the labor market and education, and whether these beliefs are sensitive to the exact formulation of he question. Table S.2 provides an overview of the treatment design in the policy experiment.

[3]We implemented this survey on a different sample than the choice experiment to avoid any confounds between the spectators' choices and elicited beliefs. We did not elicit beliefs for the single-gender luck treatments, since there is no basis for the spectators having different beliefs about the effort of the worker falling behind in these treatments.

2.3 Participants

We collected background characteristics from all participants in terms of gender, age, income, and political orientation. Table 2 presents an overview of the background characteristics of the participants by experiment and a comparison with US census data. We observe that there are only small differences in the background characteristics of the participants in the choice experiment and in the policy experiment. The samples are fairly balanced in terms of gender, with about 46% males. The median age of the participants is 46 years and the median yearly household income before taxes is around 50,000 USD. About one-third of the participants state that they would vote Republican. Overall, the samples are largely representative of the US population (+18 years old) on these dimensions, even though we note that income is more compressed in the samples than in the population at large (which may partly reflect that self-reported income was restricted at the extremes). In Tables S.3-S.6, we show that the samples are balanced across treatments.

[Table 2 about here]

3 Empirical strategy

We here outline the empirical strategy used in the main analysis of the choice experiment and the policy experiment.

In the analysis of the choice experiment, we focus on the following empirical specification:

$$u_i = \alpha + \beta_1 Malebehind_i + \beta_2 Luck_i + \beta_3 Malebehind_i \times Luck_i + \gamma \mathbf{X}_i + \epsilon_i, \quad (1)$$

where u_i is an indicator for whether spectator i has transferred nothing to the worker with no earnings or the standardized amount transferred, $Malebehind_i$ is an indicator variable for the spectator being in a treatment where a male has fallen behind, $Luck_i$ is an indicator for the spectator being in a treatment where the inequality in earnings is determined by luck, $Malebehind \times Luck_i$ is an interaction variable between the two treatment indicators, $\mathbf{X}_i$ is a vector of control variables, and ϵ_i is an error term. We report equation (1) with and without the set of control variables (gender, political affiliation, income, and age), and separately for the mixed-gender treatments and the single-gender treatments. Correspondingly, when analyzing elicited beliefs about the effort exerted by the two workers, we estimate (1) with the dependent variable being an indicator for whether the spectator agrees (somewhat or strongly) that the worker with no earnings has exerted less effort than the worker with earnings or the standardized level of agreement (strongly disagree (1) - strongly agree (5)).

In the regression analysis of equation (1), β_1 provides an estimate of the causal effect of the gender of the worker with no earnings on the spectators' decisions or on the elicited beliefs, β_2

an estimate of the causal effect of the inequality in earnings being determined by luck rather than merit, and β_3 an estimate of the interaction effect between the gender of the worker with no earnings and the source of inequality.

We study heterogeneous effects in both the spectators' decisions and in elicited beliefs for gender, political orientation, income, and the age of the participants. To illustrate, we use the following specification when studying gender differences in the mixed-gender merit treatments:

$$u_i = \alpha + \beta_1 Malebehind_i + \beta_2 M_i + \beta_3 M_i \times Malebehind_i + \epsilon_i, \tag{2}$$

where u_i is the relevant dependent variable (spectator's decision or elicited beliefs), M_i is an indicator variable for participant i being male, and $M_i \times Malebehind_i$ is an interaction variable for the participant being male and in a treatment where a male worker has no earnings. In this analysis, β_1 provides an estimate of the causal effect of the gender of the worker having no earnings for the female participants, $\beta_1 + \beta_3$ provides the corresponding estimate for the male participants, and β_3 provides an estimate of whether the causal effect differs between the female participants and the male participants. We report corresponding regressions for the other background characteristics. In the heterogeneity analysis, we only report results for the standardized dependent variable, but the findings are robust to using the corresponding indicator variable.

The analysis of the policy experiment follows the same structure as for the choice experiment, but the main regression equation is simplified because we do not manipulate why people fall behind in the policy experiment. Hence, we estimate the following regression:

$$u_i = \alpha + \beta_1 Malebehind_i + \gamma \mathbf{X}_i + \epsilon_i, \tag{3}$$

where u_i is the relevant dependent variable (policy support or elicited beliefs) and $Malebehind_i$ is an indicator variable for participant i being in a treatment where the question is about males falling behind in the labor market and education. We report equation (3) with and without the set of control variables (gender, political affiliation, income, and age). We consider two dependent variables when studying policy support: an indicator variable for whether the participant agrees (somewhat or strongly) that the government should provide support to people falling behind in the labor market and education or the standardized level of agreement (strongly disagree (1) - strongly agree (5)). Correspondingly, we consider two dependent variables when studying beliefs about effort: an indicator variable for whether the spectator agrees (somewhat or strongly) or the standardized level of agreement (strongly disagree (1) - strongly agree(5)). The two versions of the belief question give very similar results (see Figure S.4), and thus we pool them in the main analysis. In the heterogeneity analysis, we estimate a version of equation (2) with the relevant dependent variables for policy support and elicited beliefs.

In the analysis of both the choice experiment and the policy experiment, we compute p-values adjusted for multiple hypothesis testing as a robustness check of the main results. We calculate unadjusted p-values as bootstrap p-values following Davison and Hinkley (1997), and compute p-values adjusted for step-down multiple testing following the algorithm proposed by Romano and Wolf (2016). Bootstrapping is implemented with 10,000 replications. All the main findings are robust to multiple hypothesis testing, as shown in Appendix A.2.

4 Choice experiment: Results

We first provide an overview of the spectators' decisions and beliefs about effort in the choice experiment, before we turn to the regression analysis of the main treatments and the heterogeneity analysis.

4.1 Descriptive statistics

Figure 2 gives an overview of the distributions of the spectators' transfers (upper panel) and beliefs about effort (lower panel) pooled for all treatments, see Figures S.1–S.2 for the corresponding distributions by treatment.

We observe from the upper panel that the most common spectator allocations are to give nothing to the worker with no earnings (30%) and to equalize (29%). The large majority of spectators (67.2%) give more to the worker with earnings than to the worker with no earnings, while very few spectators (4%) give more to the worker with no earnings. The lower panel shows that about 40% of the participants strongly or somewhat agree that the worker with no earnings has exerted less effort than the worker ahead. The mode is neither to agree nor disagree, while about 20% somewhat or strongly disagree.

[Figure 2 about here]

4.2 Main findings

We start by discussing the regression analysis of the mixed-gender treatments (merit and luck), where the initial inequality in earnings between a male worker and a female worker is determined by the productivity of the workers or by luck. Table 3 reports the estimates of equation (1), where columns (1)–(4) report the regression analysis for the spectators' redistributive decisions and columns (5)–(8) report the regression analysis for the elicited beliefs about effort.

[Table 3 about here]

In column (1), we observe a significant effect of manipulating the gender of the worker with no earnings on the spectators' choices in the merit treatments: the share of spectators transferring nothing to the low productive worker with no earnings increases by 7.3 percentage points when this worker is a male rather than a female. In column (2), we show that the estimated treatment effect is robust to including background variables for the gender of the spectator, political orientation, income, and age, and in columns (3)–(4) that it is robust to using the standardized amount transferred as the dependent variable. The standardized amount transferred to the worker with no earnings is reduced by about 0.13 standard deviations when the low productive worker is a male rather than a female. We note that the estimated treatment effect of manipulating the gender of the worker with no earnings is as large as the estimated effect of being Republican when considering the share transferring nothing to the worker with no earnings, and two-thirds of the effect of being Republican when considering the amount transferred.

In columns (5)–(8), we find a corresponding pattern for elicited beliefs in the merit treatments. In column (5), we observe that the share of participants agreeing that the worker with no earnings has exerted less effort than the worker with earnings increases with 8.7 percentage points when the low productive worker is a male rather than a female, which is close to the estimated treatment effect on the share of participants giving nothing to the worker with no earnings (column (1)). Hence, the evidence suggests that more participants transfer nothing to low productive males with no earnings than to low productive females with no earnings in the merit treatments because they are more likely to believe that low productive males have exerted low effort than low productive females. The estimated treatment effect of manipulating the gender of the low productive worker on beliefs about effort is robust to including background characteristics (column (6)), and to using the standardized level of agreement as the dependent variable (columns (7)–(8)). We find that the level of agreement increases with about 0.22 standard deviations when the worker with no earning is a male rather than a female.

We summarize these findings as our first main result:

Result 1: *People redistribute less to a low productive male with no earnings than to a low productive female with no earnings, and people agree more that a low productive male has exerted low effort than a low productive female.*

In Figure 3, we report the estimated treatment effects on the spectators' choices and elicited beliefs by subgroup for the mixed-gender merit treatments, see Tables S.7–8 for the complete set of regression estimates. In Panel A, we observe that the estimated treatment effect on the amount transferred to the low productive worker with no earnings is negative and significant in all subgroups. Hence, across subgroups, we find that people transfer less to a low productive male with no earnings than to a low productive female with no earnings. We do not find significant

differences in the spectators' choices across subgroups. In Panel B, we report the corresponding heterogeneity analysis for beliefs about effort. We observe that the estimated treatment effect on the level of agreement is positive for all subgroups, which means that, across subgroups, we find that participants agree more that the low productive worker has exerted low effort when the low productive worker is a male rather than a female. The estimated treatment effect is significant for all subgroups, except for males and young participants. We summarize these findings as our second main result:

[Figure 3 about here]

Result 2: *In all subgroups in the population, we find less willingness to redistribute and more agreement with the low productive worker with no earnings having exerted low effort when the low productive worker is a male rather than a female.*

Table 3 further studies how the spectators' choices and elicited beliefs are affected by manipulating the source of inequality. We observe that when the initial earnings are determined by luck rather than merit, the share of spectators transferring nothing decreases significantly (columns (1)–(2)), and that the transferred amount increases significantly (columns (3)–(4)), in line with the existing evidence in the literature (Almås et al., 2020). We observe that there is a significant interaction effect between the source of the inequality and the gender of the worker with no earnings. The gender effect on the share transferring nothing and on the transferred amount is larger in the mixed-gender merit treatments than in the mixed-gender luck treatments. However, even in the mixed-gender luck treatments, we find that more people transfer nothing and that the transferred amount is smaller when a male has no earnings than when a female has no earnings.

In columns (5)–(8) in Table 3, we report the elicited beliefs for the mixed-gender luck treatments. We observe that, as expected, the share of participants agreeing that the worker with no earnings has exerted less effort than the worker with earnings decreases significantly when the earnings have been allocated randomly. However, also in the mixed-gender luck treatments, participants believe that it is more likely that a male with no earnings has exerted low effort than a female with no earnings, which implies that participants believe that it is more likely that the average male worker has exerted low effort than the average female worker (since earnings is an uninformative signal about productivity in the luck treatments). The fact the participants transfer less to a male with no earnings than to a female with no earnings also in the luck treatments is thus consistent with statistical fairness discrimination: the participants believe even in the luck treatment that it is more likely that a male with no earnings has exerted low effort than a female with no earnings. The gender effect on the share agreeing that the worker with no earnings has exerted low effort is in fact equally large in the mixed-gender luck treatments as in the mixed-gender merit

treatments.[4] Hence, the fact that the gender effect on the transfer to the worker with no earnings is lower in the luck treatments than in the merit treatment, despite beliefs about the gender difference in effort being the same, suggests that people consider it more justifiable to accept an inequality based on effort considerations in the merit treatments than in the luck treatments.

We also implemented a set of single-gender treatments (merit and luck) in the first round of data collection, to study the effect of the gender composition of the pair of workers on the spectators' choices, see Table S.9. We do not find any evidence of the gender of the two workers having a significant effect on the share transferring nothing or on the amount transferred to the worker with no earnings, both in the single-gender merit treatments and the single-gender luck treatments. Hence, it is not the case that people generally treat males and females differently, this pattern only emerges when they consider a distributive situation involving both a male worker and a female worker. Further, we also show in Table S.9 that people's beliefs about the effort of the low productive worker in the single-gender merit treatments do not depend on whether they consider two male workers or two female workers. Finally, in Table S.10, we show that male workers with no earnings receive less and are more likely to receive nothing in the mixed-gender merit treatments than in the single-gender merit treatments, while female workers with no earnings are treated in the same way in the mixed-gender merit treatments and the single-gender merit treatments. Taken together, the evidence suggests that people consider males who are less productive than females to be particularly undeserving of support, because they consider it a strong signal of a male having exerted low effort if a male is less productive than a female.

5 Policy experiment: Results

In this section, we study whether people are also more accepting of males falling behind than females falling behind when considering support for governmental interventions, and whether they believe that males to a greater extent than females are falling behind in the labor market and education due to lack of effort. We first provide an overview of the policy attitudes and beliefs, before we turn to the regression analysis of the treatment effects and the heterogeneity analysis.

5.1 Descriptive statistics

Figure 4 gives an overview of the distributions of the policy attitudes (upper panel) and beliefs about effort (lower panel) pooled for whether the question is about males or females falling behind, see Figure S.3 for the corresponding distributions by treatment.

[4] We observe from columns (7)-(8) that the treatment effect on the standardized level of agreement is even larger in the luck treatments than in the merit treatments. It reflects that a much larger share of people strongly disagree that a female with no earnings in the luck treatment have exerted low effort than a male with no earnings, see Figure S.2.

We observe from the upper panel that there is disagreement about this type of government interventions: 27.0 percent of the participants strongly or somewhat disagree with these governmental policies, while 48.0 percent strongly or somewhat agree. In the lower panel, we observe that there is a corresponding disagreement in the beliefs about whether people are falling behind in the labor market and education largely due to their lack of effort: 40.0 percent strongly or somewhat agree, while 27.5 percent strongly or somewhat disagree.

[Figure 2 about here]

5.2 Main findings

We start by discussing the regression analysis of how the gender of the person falling behind affects policy attitudes and beliefs about effort. Table 4 reports the estimates of equation (3), where columns (1)–(4) report the regression analysis for the participants' policy attitudes and columns (5)–(8) report the regression analysis for the elicited beliefs about effort.

[Table 4 about here]

In column (1), we observe a significant effect of manipulating the gender of people falling behind on support for government interventions: the share of participants agreeing with government interventions is reduced by 11.9 percentage points when those falling behind are males rather than females. In column (2), we show that the estimated treatment effect is robust to including background variables for the gender of the participant, political orientation, income, and age, and in columns (3)–(4) that it is robust to using the standardized level of agreement as the dependent variable. The standardized level of agreement is reduced by about 0.27 standard deviations when the question is about the government providing support to males falling behind rather than to females falling behind, which is an even larger gender-effect than we observe in the redistributive decisions in the choice experiment (see Table 3). We note that the estimated treatment effect of manipulating the gender of those falling behind on the support for government interventions is about half of the estimated effect of being Republican, both in terms of the share agreeing with government interventions and the standardized level of agreement.

In columns (5)–(8), we find a corresponding pattern for elicited beliefs about effort. In column (5), we observe that the share of participants agreeing that those falling behind have exerted low effort increases with 11.8 percentage points when those falling behind are males rather females, which is close to the estimated reduction in the share of participants agreeing with government interventions when those falling behind are males rather females (column (1)). Hence, as in the choice experiment, we observe that the estimated effect on beliefs about effort is comparable to the estimated effect on the acceptability of the inequality. The evidence thus suggests that fewer

participants support government interventions when males fall behind than when females fall behind because they are more likely to believe that males falling behind have exerted low effort than females falling behind. The estimated treatment effect of manipulating the gender of those falling behind on beliefs about effort is robust to including background characteristics (column (6)), and to using the standardized level of agreement as the dependent variable (columns (7)–(8)). We find that the standardized level of agreement increases with about 0.46 standard deviations when those falling behind are males rather than females, which is even greater than the effect on beliefs of effort when manipulating the gender in the choice experiment. The estimated treatment effect of manipulating the gender of those falling behind on beliefs about effort is comparable to the estimated effect of being Republican, both in terms of the share agreeing to those falling behind having exerted low effort and the standardized level of agreement. Taken together, we state the following result:

Result 3: *People agree less to the government providing support to males falling behind in the labor market and education than to females falling behind, and people agree more that males falling behind have exerted low effort than females falling behind.*

[Figure 5 about here]

In Figure 5, we report the estimated treatment effects on policy attitudes and beliefs about effort by subgroup, see Tables S.11–S.12 for the complete set of regression estimates. In Panel A, we observe that the estimated treatment effect on the support for government interventions supporting those who fall behind is negative and significant in all subgroups. Hence, across subgroups, we find that people agree less to government interventions supporting when a male rather than a female is falling behind. We find that the gender-effect is smaller for males compared to females, Republicans compared to non-Republicans, and younger compared to older participants. In Panel B, we report the corresponding heterogeneity analysis for beliefs about effort. We observe that the estimated treatment effect on the level of agreement is positive for all subgroups, which means that, across subgroups, we find that participants agree more that males falling behind have exerted low effort than females falling behind. We find the largest difference in the estimated treatment effect on beliefs about effort between young and older participants, with the older participants differentiating significantly more their beliefs about the effort based on the gender of those falling behind. We summarize the heterogeneity analysis in the following result:

Result 4: *In all subgroups in the population, we find less agreement with the government providing support and more agreement with those falling behind in the labor market and education having exerted low effort when considering males falling behind compared to females falling behind.*

6 Concluding remarks

In recent decades, disadvantaged males have experienced a downward trajectory in the labor market and education in many countries, which partly reflects structural changes and socioeconomic developments (Autor and Wasserman, 2013). How society responds to this development may depend critically on how we perceive males falling behind. We have shown in a set of large-scale experimental studies conducted with general population samples from the US that people are more accepting of males falling behind than females falling behind, and we provide evidence suggesting that the greater acceptance of males falling behind reflects that people consider males falling behind to be less deserving of support than females falling behind. We find these patterns both in a large-scale choice experiment studying redistribution in a controlled work environment, and in a large-scale policy experiment studying people's agreement to policy interventions supporting those who fall behind in the labor market and education. We argue that the observed pattern may be seen as reflecting statistical fairness discrimination based on a meritocratic fairness view: people are more accepting of males falling behind than females falling behind because they are more likely to believe that males fall behind due to low effort than females.

Statistical fairness discrimination may shed light on a wide range of discriminatory behavior. To illustrate, consider the accumulated non-experimental evidence suggesting, albeit with some exceptions, that teachers' grade setting is biased against males in a range of subjects in kindergarten, primary school, high school, and higher education (Breda and Ly, 2015; Cornwell, Mustard, and Van Parys, 2013; Falch and Naper, 2013; Lavy, 2008; Lavy and Sand, 2018; Lindahl, 2016; Terrier, 2020). Such behavior may reflect statistical fairness discrimination, where teachers may be more likely to interpret poor performance of male students to reflect lack of effort than poor performance of female students. It would be of great interest in future research to explore statistical fairness discrimination in field studies in education and at the workplace, and establish more evidence on its relevance for explaining gender discriminatory behavior. It would also be important to further our understanding of why people view lack of effort to be more important for males falling behind than for females falling behind, and to understand how this relates to stereotypes that people have about males and females (Alan, Ertac, and Mumcu, 2018; Bordalo, Coffman, Gennaioli, and Shleifer., 2019).

We have focused on how statistical fairness discrimination may disadvantage males falling behind, but this mechanism may in other contexts disadvantage females. For example, affirmative action for females may make people less supportive of females falling behind, if affirmative action policies make people believe that females who fall behind have exerted low effort. More broadly, statistical fairness discrimination may also contribute to explain other types of discrimination, including discrimination of ethnic minorities and immigrants (Cettolin and Suetens, 2019), where

people may hold stereotypes about the effort exerted by these groups.

We have suggested an experimental paradigm for studying statistical fairness discrimination that is portable to a wide range of different contexts. We hope that future research will use this approach to deepen our understanding of how statistical fairness discrimination shape behavior and policies. In an era where politics has become "a personal responsibility crusade" (Hacker, 2006), we need to carefully examine how we relate to people who are struggling in society (Moffitt, 2015).

Acknowledgements. The study was organized by FAIR at NHH Norwegian School of Economics. We are grateful to Daniel Benjamin, Pedro Rey Biel, Caroline Bonn, Gary Charness, Stefano DellaVigna, Catherine Eckel, Christine L. Exley, Uri Gneezy, Ingar Haaland, Alex Imas, Siri Isaksson, Elisabeth Ivarsflaten, Nanjundi Karthick Krishnan, James Konow, Erin Krupka, Erzo F.P. Luttmer, Axel Ockenfels, Anders Poulsen, Odile Poulsen, Lise Rakner, Imran Rasul, Charlotte Ringdal, Kjell Salvanes, Marta Serra-Garcia, Ingrid Hoem Sjursen, Charles Sprenger, Eirik Strømland, Matthias Sutter, Jonas Tungodden, Lise Vesterlund and numerous seminar participants for valuable comments and suggestions and to Sebastian Fest for valuable research assistance. We are also thankful for the financial support from the NORFACE (New Opportunities for Research Funding Agency Cooperation in Europe) Network, The Centre for Ethics and Economics at the Norwegian School of Economics and the Research Council of Norway through its Centres of Excellence Scheme, Centre for Experimental Research on Fairness, Inequality and Rationality (FAIR) project (no. 262675). The study was pre-registered in the AEA RCT Registry: AEARCTR-0000853, AEARCTR-0001027 and AEARCTR-0005610.

References

Alan, Sule, Seda Ertac, and Ipek Mumcu (2018). "Gender stereotypes in the classroom and effects on achievement," *Review of Economics and Statistics*, 100(5): 876–890.

Almås, Ingvild, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø. Sørensen, and Bertil Tungodden (2016). "What explains the gender gap in college track dropout? Experimental and administrative evidence," *American Economic Review: Papers & Proceedings*, 106(5): 296–302.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden (2020). "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?" *Journal of Political Economy*, 128(5): 1753–1788.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2019). "Family disadvantage and the gender gap in behavioral and educational outcomes," *American Economic Journal: Applied Economics*, 11: 338–381.

Autor, David and Melanie Wasserman (2013). "Wayward sons: The emerging gender gap in labor markets and education," Technical report, Third Way.

Babcock, Linda, Maria P. Recalde, Lise Vesterlund, and Laurie Weingart (2017). "Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability," *American Economic Review*, 107(3): 714–747.

Bertrand, Marianne and Esther Duflo (2017). *Handbook of economic field experiments*, volume 1, chapter Field experiments on discrimination, pp. 309–393.

Bertrand, Marianne and Jessica Pan (2013). "The trouble with boys: Social influences and the gender gap in disruptive behavior," *American Economic Journal: Applied Economics*, 5(1): 32–64.

Binder, Ariel J. and John Bound (2019). "The declining labor market prospects of less-educated men," *Journal of Economic Perspectives*, 33: 163–190.

Blau, Francine D. and M. Kahn, Lawrence (2013). "Female labor supply: Why is the United States falling behind?" *American Economic Review: Papers and Proceedings*, 103(3): 251–256.

Bohren, Aislinn J., Alex Imas, and Michael Rosenberg (2019). "The dynamics of discrimination: Theory and evidence," *American Economic Review*, 109: 3395–3436.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope (forthcoming). "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*.

Bolton, Gary E. and Axel Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, 90(1): 166–193.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. (2019). "Beliefs about gender," *American Economic Review*, 109(3): 739–773.

Bortolotti, Stefania, Ivan Soraperra, Matthias Sutter, and Claudia Zoller (2017). "Too lucky to be true. Fairness views under the shadow of cheating," CESifo Working Paper Series No. 6563.

Breda, Thomas and Son Thierry Ly (2015). "Professors in core science fields are not always biased against women: Evidence from France," *American Economic Journal: Applied Economics*, 7(4): 53–75.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden (2007). "The pluralism of fairness ideals: An experimental approach," *American Economic Review*, 97(3): 818–827.

Cappelen, Alexander W., Ranveig Falch, and Bertil Tungodden (2020). *Handbook of Labor, Human Resources and Population Economics*, chapter Fair and Unfair Income Inequality, Springer, Cham., pp. 1–25.

Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden (2013). "Just luck: An experimental study of risk taking and fairness," *American Economic Review*, 103(3): 1398–1413.

Castillo, Marco, Ragan Petrie, Maximo Torero, and Lise Vesterlund (2013). "Gender differences in bargaining outcomes: A field experiment on discrimination," *Journal of Public Economics*, 99(C): 35–48.

Cettolin, Elena and Sigrid Suetens (2019). "Return on trust is lower for immigrants," *The Economic Journal*, 129: 1992–2009.

Coffman, Katherine B., Christine L. Exley, and Muriel Niederle (2021). "The role of beliefs in driving gender discrimination," *Management Science*, 67(6): 3321–3984.

Coile, Courtney C. and Mark G. Duggan (2019). "When labor's lost: Health, family life, incarceration, and education in a time of declining economic opportunity for low-skilled men," *Journal of Economic Perspectives*, 33: 191–210.

Cornwell, Christopher, David B. Mustard, and Jessica Van Parys (2013). "Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school," *Journal of Human Resources*, 48(1): 236–264.

Council of Economic Advisers (2016). "The long-term decline in US prime-age male labour force participation," Technical report, Executive Office of the President.

Croson, Rachel and Uri Gneezy (2009). "Gender differences in preferences," *Journal of Economic Literature*, 47(2): 1–27.

Davison, A.C. and D. V. Hinkley (1997). *Bootstrap methods and their application*, Cambridge University Press.

Eckel, Catherine C. and Philip J. Grossman (2001). "Chivalry and solidarity in ultimatum games," *Economic Inquiry*, 39(2): 171–188.

Eckel, Catherine C. and Philip J. Grossman (2008). *Differences in the economic decisions of men and women: Experimental evidence*, volume 1, chapter 57, Handbook of experimental economic results, Elsevier, 1st edition, pp. 509–519.

Engel, Christoph (2011). "Dictator Games: A Meta Study," *Experimental Economics*, 14(4): 583–610.

Exley, Christine L., Muriel Niederle, and Lise Vesterlund (2020). "Knowing when to ask: The cost of leaning-in," *Journal of Political Economy*, 128(3): 816–854.

Falch, Ranveig (2022). "How do people trade off resources between quick and slow learners?" *European Economic Review*, 150(104285): 1–21.

Falch, Torberg and Linn Renée Naper (2013). "Educational evaluation schemes and gender gaps in student achievement," *Economics of Education Review*, 36(C): 12–25.

Fehr, Ernst and Klaus M. Schmidt (1999). "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics*, 114(3): 817–868.

Goldin, Claudia and Cecilia Rouse (2000). "Orchestrating impartiality: The impact of "blind" auditions on female musicians," *American Economic Review*, 90(4): 715–741.

Hacker, Jacob S. (2006). *The great risk shift*, Oxford University Press.

Hvidberg, Kristoffer B., Claus T. Kreiner, and Stefanie Stantcheva (forthcoming). "Social positions and fairness views on inequality," *Review of Economic Studies*.

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

Krueger, Alan B. (2017). "Where have all the workers gone?" *Brookings Papers on Economic Activity*, 48: 1–87.

Lavy, Victor (2008). "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment," *Journal of Public Economics*, 92(10-11): 2083–2105.

Lavy, Victor and Edith Sand (2018). "On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases," *Journal of Public Economics*, 167: 263–279.

Lindahl, Erica (2016). "Are teacher assessments biased? Evidence from Sweden," *Education Economics*, 24(2): 224–238.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz (2018). "Gender bias in teaching evaluations," *Journal of the European Economic Association*.

Moffitt, Robert A. (2015). "The deserving poor, the family, and the U.S. welfare system," *Demography*, 52(3): 729–749.

OECD (2015). "The abc of gender equality in education: Aptitude, behaviour, confidence," Technical report, PISA, OECD Publishing.

OECD (2022). "Education at a glance 2022: OECD indicators," Technical report, OECD, Publishing Paris.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2014). "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences*, 111(12): 4403–4408.

Reynolds, Tania, Chuck Howard, Hallgeir Sjåstad, Luke Zhu, Tyler G. Okimoto, Roy F. Baumeister, Karl Aquino, and JongHan Kim (2020). "Man up and take it: Gender bias in moral typecasting," *Organizational Behavior and Human Decision Processes*, 161: 120–141, organizational Behavior and Human Decision Processes.

Romano, Joseph P. and Michael Wolf (2016). "Efficient computation of adjusted p-values for resampling-based stepdown multiple testing," *Statistics & Probability Letters*, 113(1): 38–40.

Rosin, Hanna (2012). *The end of men and the rise of women*, New York: Riverhead Books.

Solnick, Sara J. (2001). "Gender differences in the ultimatum game," *Economic Inquiry*, 39(2): 189–200.

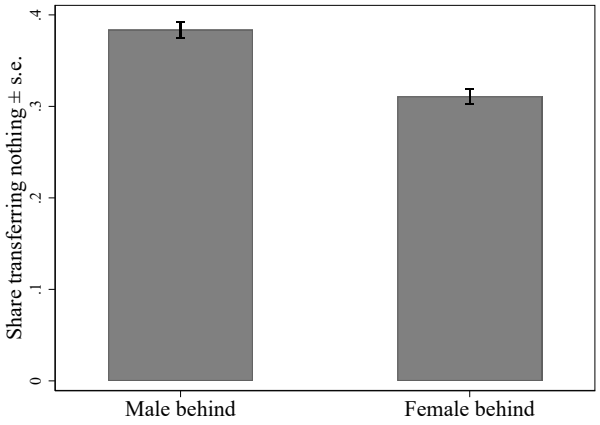
Terrier, Camille (2020). "Boys lag behind: How teachers' gender biases affect student achievement," *Economics of Education Review*, 77(101981).

U.S. Bureau of Labor Statistics (2022). "Median earnings for women in 2021 were 83.1 percent of the median for men," Bureau of Labor Statistics, U.S. Department of Labor, The Economics Daily, visited December 14, 2022.

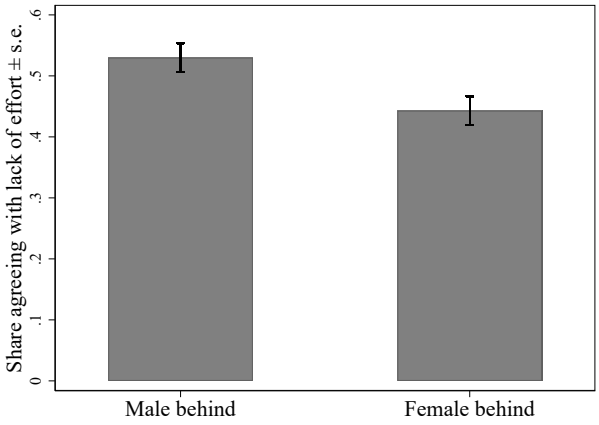
Williams, Wendy M. and Stephen J. Ceci (2015). "National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track," *Proceedings of the National Academy of Sciences*, 112(17): 5360–5365.

Figure 1: Main findings

Choice experiment

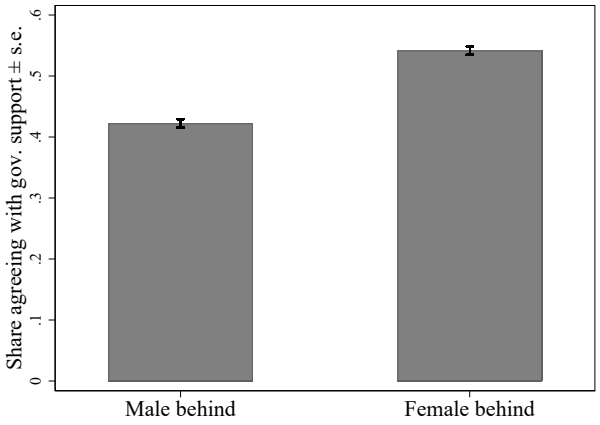


Panel A: Share transferring nothing

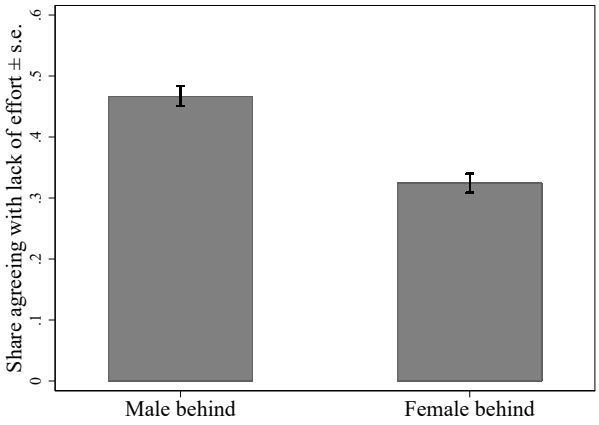


Panel B: Share agreeing with lack of effort

Policy experiment



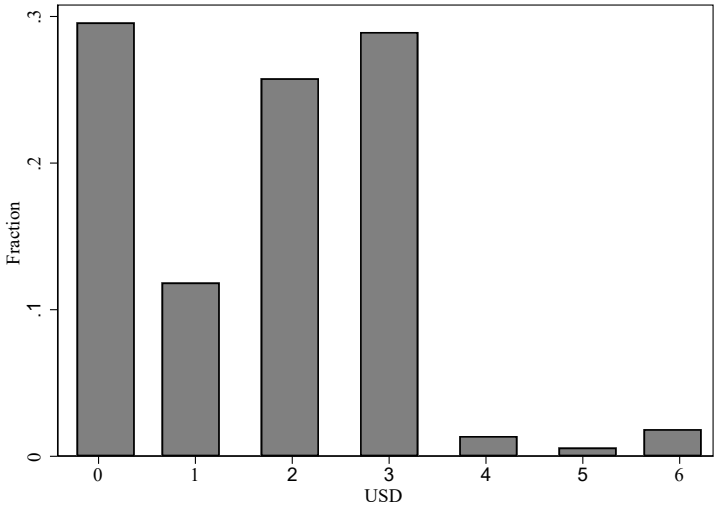
Panel C: Share agreeing with government support



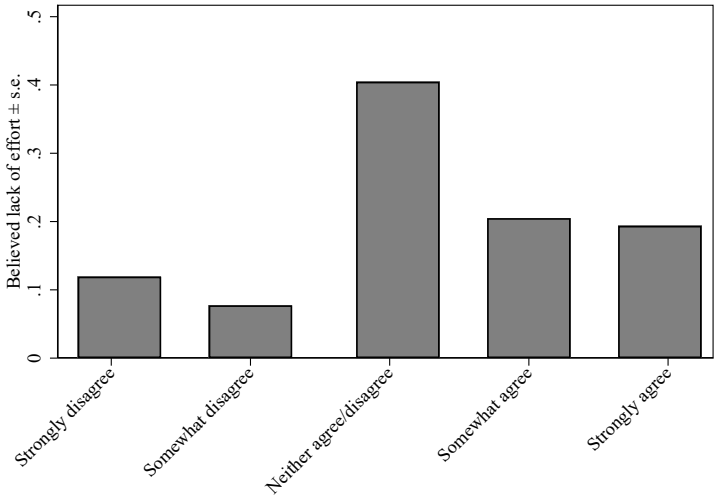
Panel D: Share agreeing with lack of effort

Note: Panel A shows the share of participants transferring nothing to the male falling behind and the female falling behind in the mixed-gender merit treatments. Panel B shows the share of participants who strongly or somewhat agree with the statement "I expect that the less productive man (woman) exerted less effort on the assignment than the more productive woman (man)." Panel C shows the share of participants who strongly or somewhat agree with the statement "It is very important that the government provides support to males (females) who fall behind in education and in the labor market." Panel D shows the share of participants who strongly or somewhat agree with the statement "When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort" or the statement "Males (Females) falling behind in education and in the labor market have exerted low effort." The standard errors are indicated by the bars.

Figure 2: Choice experiment



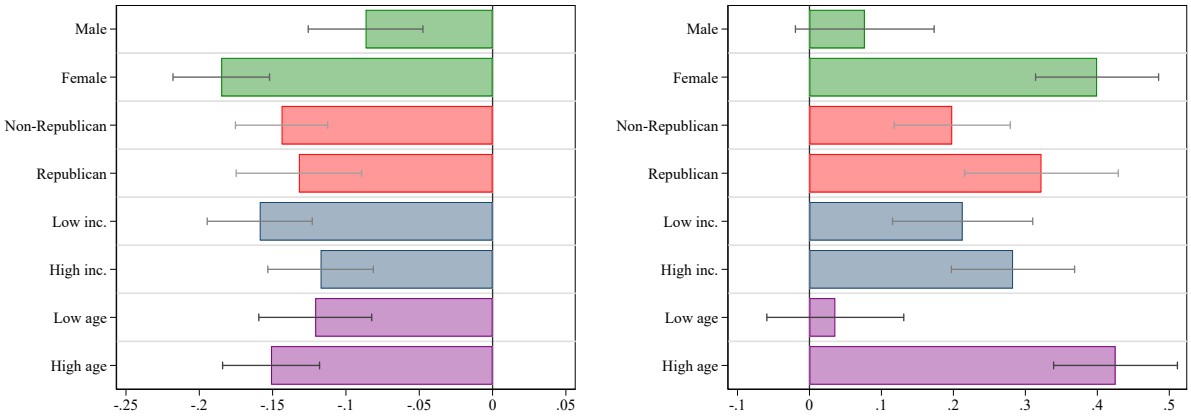
Panel A: Transfer to worker with no earnings



Panel B: Agreement lack of effort

Note: Panel A shows the distribution of transfers (in USD) to the worker falling behind, pooled for all treatments. Panel B shows the extent of agreement with the worker falling behind having exerted less effort than the worker ahead, pooled for all treatments.

Figure 3: Choice experiment, heterogeneity, mixed-gender merit

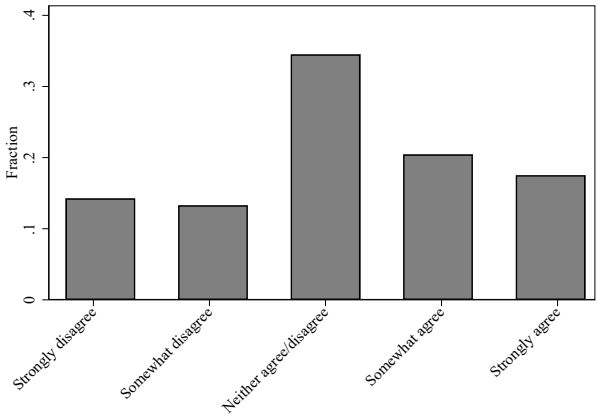


A: Amount to worker behind (std)

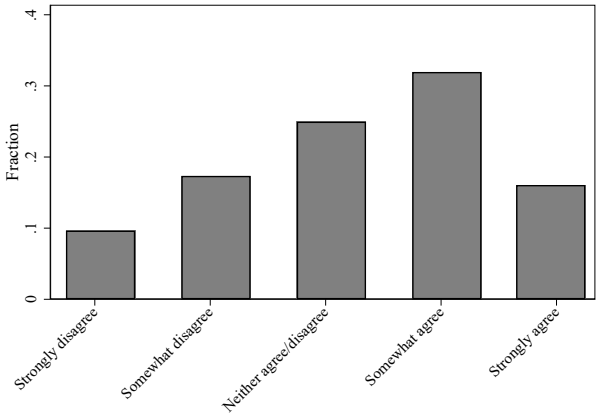
B: Agreement lack of effort (std)

Note: Panel A reports, by subgroup, the estimated treatment effect for the standardized amount transferred to the low productive worker with no earnings in the mixed-gender merit treatments. Panel B reports, by subgroup, the estimated treatment effect for the standardized level of agreement with the low productive worker having exerted low effort in the mixed-gender merit treatments. The estimates are from population-weighted linear regressions as specified in equation (2), where we run separate regressions for the background variables "Male participant," "Republican," "Low income," and "Low age," as defined in Table 3. Included in each regression is the relevant background variable and its interaction with "Male behind." The regressions underlying Panel A also include indicators for the round of the study in which the spectator took part. Robust standard errors are indicated by bars.

Figure 4: Policy experiment: Level of agreement



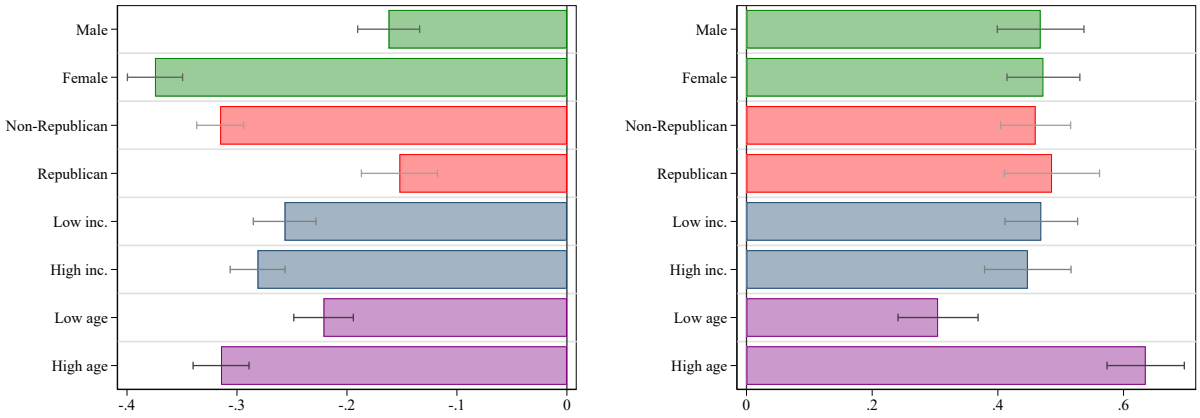
Panel A: Agreement with support



Panel B: Agreement lack of effort

Note: Panel A shows the distribution of agreement with the statement "It is very important that the government provides support to males (females) who fall behind in education and in the labor market." Panel B shows the distribution of agreement with the statement "When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort" or the statement "Males (Females) falling behind in education and in the labor market have exerted low effort." In both panels, the level of agreement is measured on a scale from strongly disagree (1) to strongly agree (5).

Figure 5: Policy experiment, heterogeneity



A: Agreement government support (std)

B: Agreement lack of effort (std)

Note: Panel A reports, by subgroup, the estimated treatment effect for the standardized level of agreement with the government providing support for people falling behind in education and in the labor market. Panel B reports, by subgroup, the estimated treatment effect for the standardized level of agreement with people falling behind in the labor market and education having exerted low effort. The estimates are from population-weighted linear regressions as specified in equation (2), where we run separate regressions for the background variables "Male participant," "Republican," "Low income," and "Low age," as defined in Table 3. Included in each regression is the relevant background variable and its interaction with "Male behind". The regressions underlying Panel B also include indicators for the round of the study in which the spectator took part. Robust standard errors are indicated by bars.

Table 1: The stages of the choice experiment: Spectators and workers

1. **Work stage:** Workers complete an assignment.
2. **Earnings stage:** Workers randomly matched in pairs. Assigned initial earnings according to treatment.
3. **Redistribution stage:** Each spectator decides for one pair of workers whether and how much to redistribute.
4. **Payment stage:** Workers in the pair paid according to the decision of the spectator.

Table 2: Descriptive statistics

	Choice experiments	Policy experiments	US
Male (share)	0.461	0.463	0.487
Age (year)			
Median	46	46	47
p10	24	23	23
p90	69	69	73
Income (USD)			
Median	57,500	48,000	63,600
p10	12,500	17,500	12,800
p90	137,500	117,900	179,600
Republican (share)	0.359	0.313	0.280
N	17,521	13,264	

Note: The table reports the descriptive statistics for the samples used in the choice experiment and the policy experiment, as well as for the US population. Samples (self-reported): the income variable is yearly household income in USD before taxes (inflation adjusted to 2020), given in standard categories where we impute the midpoint in each group to calculate the average. For the highest and lowest income groups, we impute 1.5 times the lower boundary and 0.5 times the higher boundary, respectively (see Section B.5 in the supplementary material for a list of income categories). The participants who did not know or did not want to state their income are not included in the descriptives on income (148/17,521 and 371/13,264 observations, respectively). A participant is classified as Republican if he or she would have voted for the Republican party if there was an election tomorrow. US population: the share of males and the median age (+18) in the US are estimates from the US Census Bureau, American Community Survey (2018 and 2019) (https://www.census.gov/ and https://www.census.gov/data/). The income data are based on 2018 estimates from the US Census Bureau, American Community Survey (inflation adjusted to 2020). Political affiliation is from Gallup (http://news.gallup.com/poll/).

Table 3: Choice experiment

	Spectator choice				Effort beliefs			
	Nothing to worker behind		Amount to worker behind (std)		Agree low effort		Level of agreement (std)	
Male behind	0.073***	0.074***	-0.129***	-0.130***	0.087***	0.089***	0.216***	0.217***
	(0.012)	(0.012)	(0.024)	(0.024)	(0.033)	(0.033)	(0.057)	(0.057)
Luck	-0.079***	-0.077***	0.480***	0.476***	-0.271***	-0.261***	-0.816***	-0.796***
	(0.012)	(0.012)	(0.026)	(0.026)	(0.029)	(0.029)	(0.064)	(0.062)
Luck × Male behind	-0.040**	-0.042**	0.061*	0.067*	0.011	0.003	0.328***	0.314***
	(0.017)	(0.017)	(0.037)	(0.036)	(0.043)	(0.042)	(0.086)	(0.084)
Male participant		0.039***		-0.057***		0.034		0.094**
		(0.009)		(0.018)		(0.022)		(0.044)
Republican		0.087***		-0.189***		0.015		0.036
		(0.009)		(0.019)		(0.022)		(0.044)
Low income		-0.005		0.054***		-0.040*		-0.050
		(0.009)		(0.018)		(0.022)		(0.043)
Low age		-0.029***		0.038**		0.178***		0.365***
		(0.009)		(0.018)		(0.022)		(0.043)
Constant	0.310***	0.276***	-0.168***	-0.118***	0.443***	0.356***	0.215***	0.009
	(0.009)	(0.012)	(0.018)	(0.024)	(0.023)	(0.030)	(0.042)	(0.056)
Male behind (luck)	0.033***	0.031***	-0.067**	-0.063**	0.098***	0.092***	0.544***	0.531***
	(0.012)	(0.012)	(0.028)	(0.028)	(0.027)	(0.026)	(0.064)	(0.062)
Observations	13,495	13,495	13,495	13,495	1,998	1,998	1,998	1,998
R^2	0.016	0.028	0.069	0.081	0.087	0.123	0.150	0.185

Note: The table reports population-weighted linear regressions where the dependent variable is an indicator variable for the spectator transferring nothing to the worker who is falling behind (columns (1)–(2)), the standardized amount transferred to the worker who is falling behind (columns (3)–(4)), an indicator variable for the participant strongly or somewhat agreeing with the worker falling behind having exerted less effort than the worker ahead, and the standardized level of agreement with the worker falling behind having exerted less effort than the worker ahead (columns (7)–(8)). The standardized variables are established on the basis of the responses by the participants in the respective mixed-gender treatments. "Male behind" is an indicator variable for the participant being in the treatment where the male is falling behind. "Male participant" is an indicator variable for being male. "Republican" is an indicator variable for voting Republican. "Low income" is an indicator variable for having an income below the median income per year in the respective sample. "Low age" is an indicator variable for being below 46 years old (which is the median age in the pooled sample). "Luck" is an indicator variable for the participant being in a treatment where luck is the source of inequality. "Luck × Male behind" is an interaction between "Luck" and "Male behind". The regressions in columns (1)–(4) include indicator variables for the round of the study in which the spectator took part (not reported). Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Policy experiment

	Policy support				Effort beliefs			
	Agree important to support		Level of agreement (std)		Agree low effort		Level of agreement (std)	
Male behind	-0.119***	-0.117***	-0.272***	-0.266***	0.141***	0.141***	0.462***	0.463***
	(0.009)	(0.009)	(0.019)	(0.018)	(0.022)	(0.021)	(0.046)	(0.043)
Male participant		0.017*		0.013		0.118***		0.203***
		(0.009)		(0.018)		(0.022)		(0.044)
Republican		-0.199***		-0.506***		0.130***		0.321***
		(0.010)		(0.020)		(0.023)		(0.046)
Low income		0.031***		0.115***		-0.112***		-0.217***
		(0.009)		(0.018)		(0.022)		(0.044)
Low age		0.124***		0.285***		0.191***		0.371***
		(0.009)		(0.018)		(0.022)		(0.043)
Constant	0.542***	0.523***	0.136***	0.102***	0.270***	0.132***	-0.230***	-0.507***
	(0.007)	(0.010)	(0.013)	(0.020)	(0.019)	(0.027)	(0.042)	(0.060)
Observations	11,209	11,209	11,209	11,209	2,054	2,054	2,054	2,054
R^2	0.014	0.071	0.018	0.106	0.033	0.119	0.053	0.136

Note: The table reports population-weighted linear regressions where the dependent variable is an indicator variable for the participant strongly or somewhat agree with "It is very important that the government provides support to females (males) who fall behind in education and in the labor market," the standardized level of agreement with the same statement (columns (3)–(4)), an indicator variable for the participant strongly or somewhat agreeing with the statement "When females (males) fall behind in education and in the labor market, it largely reflects their lack of effort" or the statement "Females (Males) falling behind in education and in the labor market have exerted low effort," and the standardized level of agreement with the same statements (columns (7)–(8)). The standardized variables are established on the basis of the responses by the participants in the respective treatments. "Male behind," "Male participant", "Republican", "Low income," and "Low age" are defined in Table 3. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix for Experimental Evidence on the Acceptance of Males Falling Behind

Alexander W. Cappelen,[1†] Ranveig Falch,[1] Bertil Tungodden[1*]

[1]FAIR, Department of Economics, NHH Norwegian School of Economics, Norway

[†]All authors contributed equally to this work.

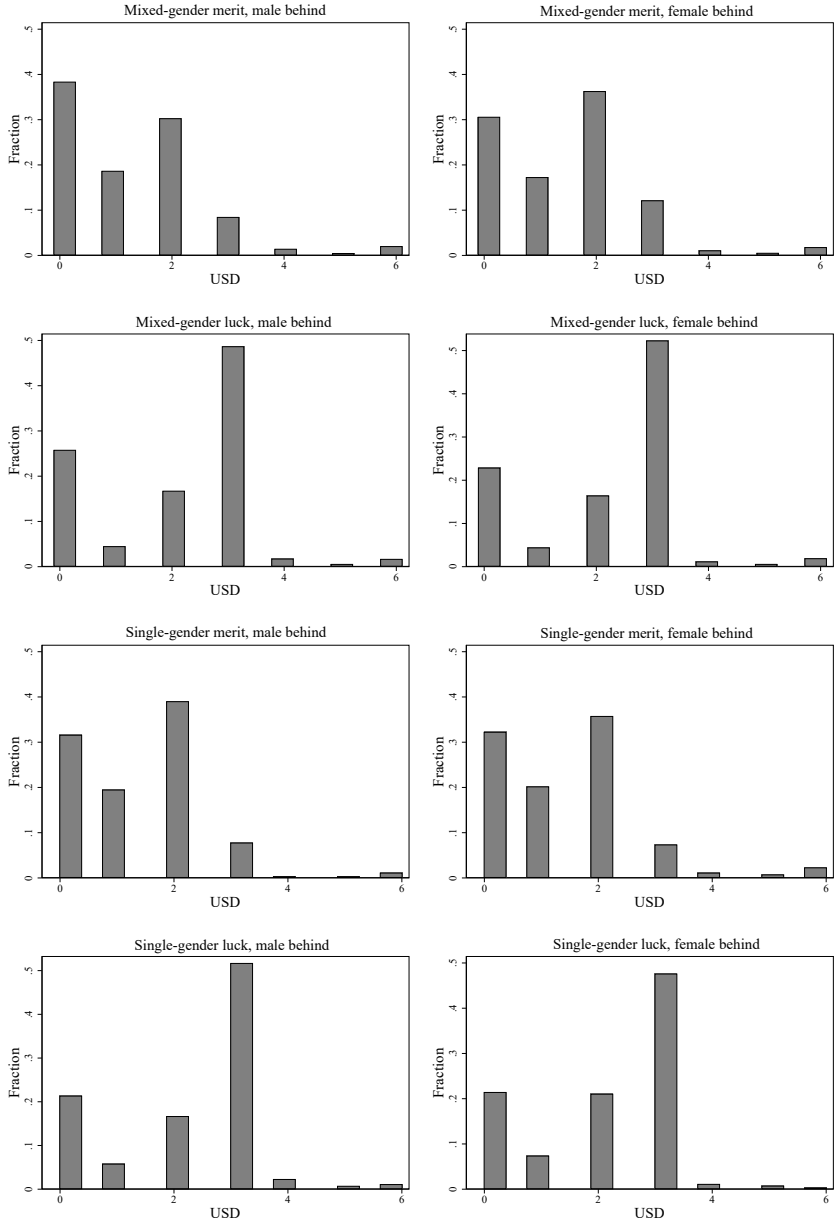
[*]To whom correspondence should be addressed; E-mail: bertil.tungodden@nhh.no.

Contents

A Online Appendix: Additional analysis

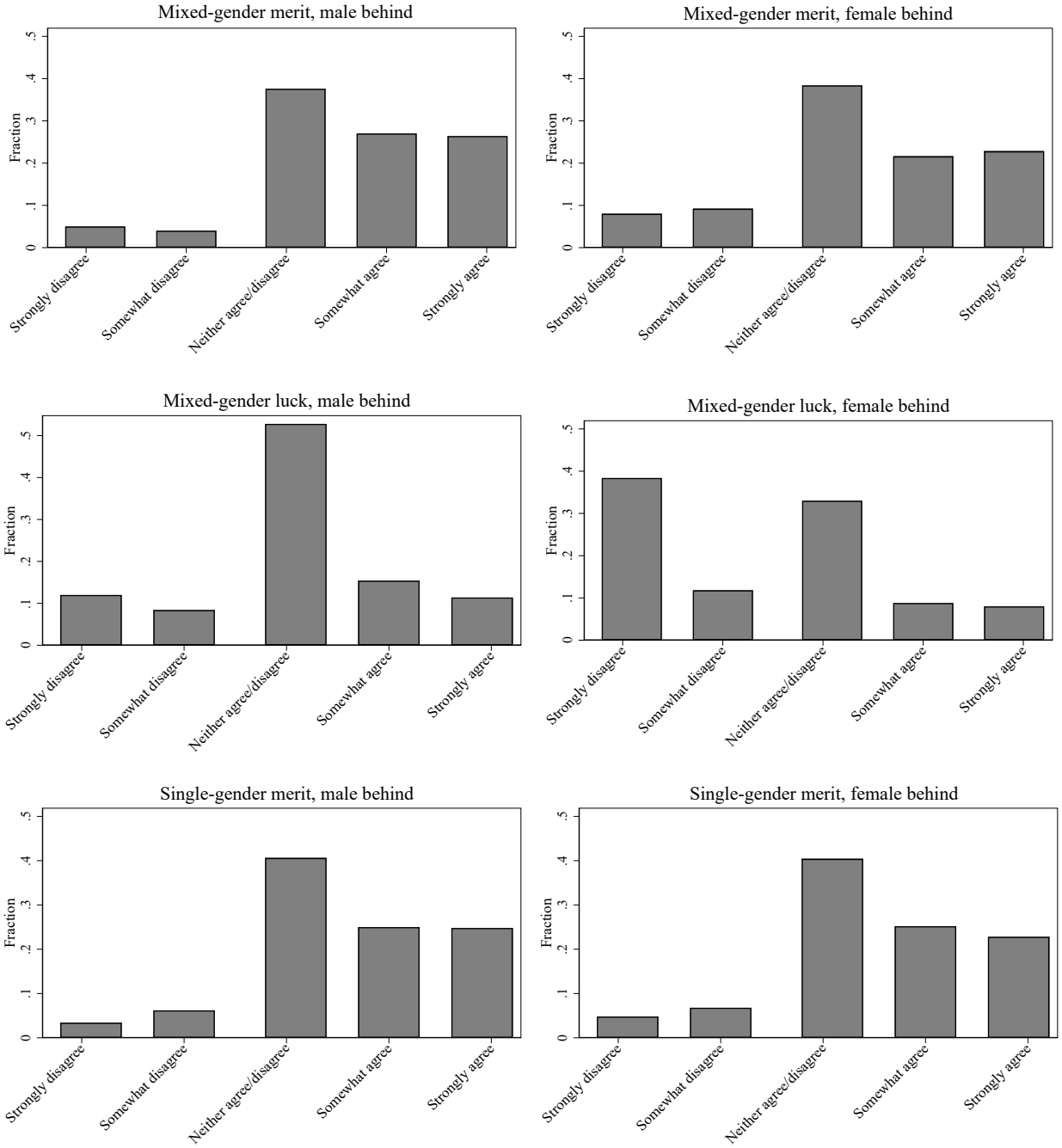
A.1 Figures and tables

Figure S.1: Choice experiment: Distribution of transfers to worker behind by treatment



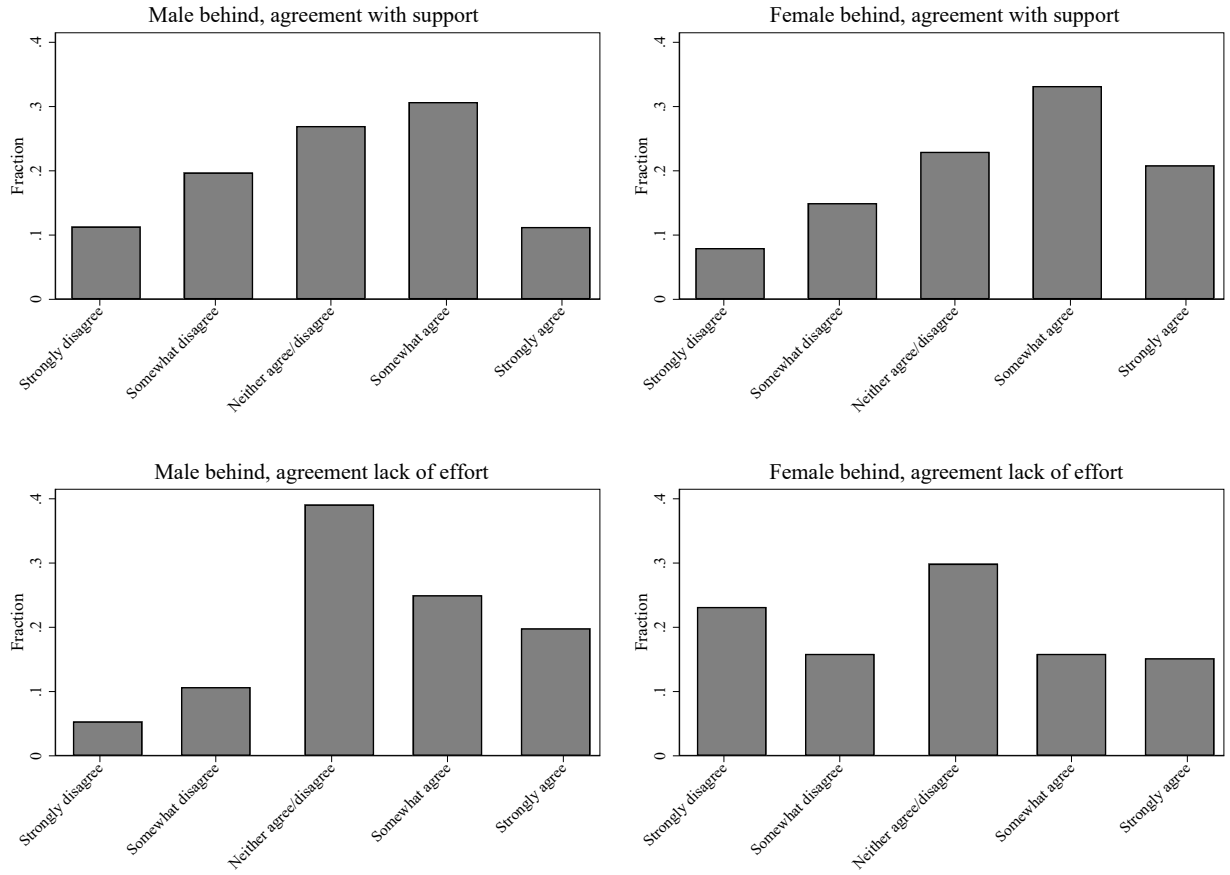
Note: The figure shows the distribution of transfers in USD to the worker who is behind by treatment.

Figure S.2: Choice experiment: Beliefs about lack of effort by treatment



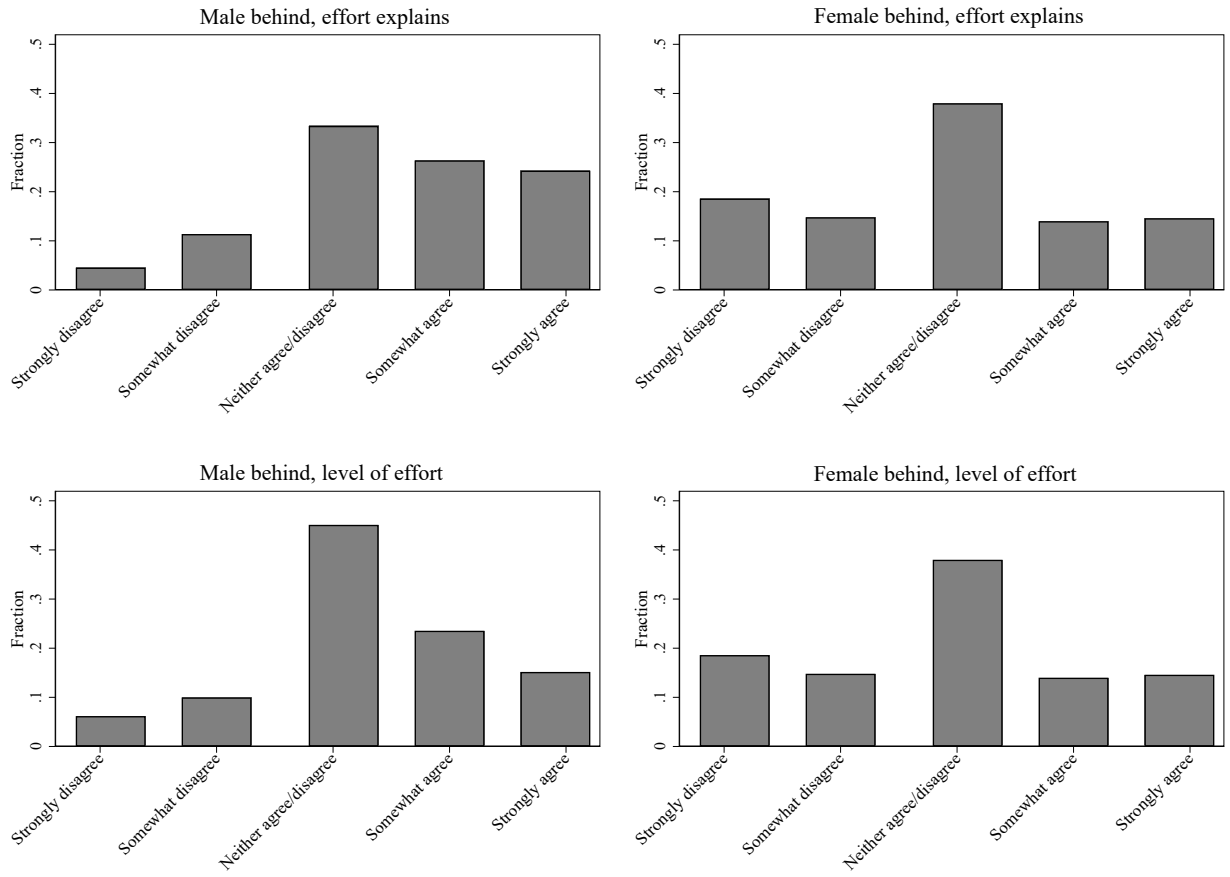
Note: The figure shows the distribution of agreement with the worker falling behind having exerted less effort than the worker ahead, by treatment.

Figure S.3: Policy experiment: Level of agreement by treatment



Note: The upper panels show the distribution of agreement with the statement "It is very important that the government provides support to males (females) who fall behind in education and in the labor market," by treatment. The lower panels show the distribution of agreement with the statement "When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort," or the statement "Males (Females) falling behind in education and in the labor market have exerted low effort," by treatment.

Figure S.4: Policy experiment: Beliefs about lack of effort by treatment and question



Note: The upper panels show the distribution of agreement with the statement "When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort," by treatment. The lower panels show the distribution of agreement with the statement "Males (Females) falling behind in education and in the labor market have exerted low effort," by treatment.

Table S.1: Choice experiment: Treatments

Spectator decisions

	Mixed-gender		Single-gender	
	Female behind	Male behind	Female behind	Male behind
MERIT	N=3,645	N=3,598	N=257	N=256
LUCK	N=3,120	N=3,132	N=256	N=257

Effort beliefs

	Mixed-gender		Single-gender	
	Female behind	Male behind	Female behind	Male behind
MERIT	N=500	N=500	N=500	N=500
LUCK	N=500	N=500		

Note: The upper panel of the table provides an overview of the eight treatments in the choice experiment and the number of spectators in each treatment. We collected the data in three rounds. First, we recruited 2,052 participants who were randomly allocated to one of eight treatments. Second, we recruited 1,050 participants who were randomly allocated to one of the two mixed-gender merit treatments, and third, we recruited 11,419 participants who were randomly allocated to one of the four mixed-gender treatments. In the first and second round, spectators were matched uniquely with a pair of workers who would be paid in accordance with their decision; in the third round, participants were informed that one out of ten spectator decisions would be randomly selected and implemented for a pair of workers. The lower panel of the table provides an overview of the independent study in which we elicited beliefs about effort in the choice experiment, by treatment. In total, we had 3,000 participants who were randomly allocated to one of six treatments.

Table S.2: Policy experiment: Treatments

	Mixed-gender	
	Female behind	Male behind
Policy preferences	N=5,608	N=5,601
Effort beliefs	N=1,027	N=1,028

Note: The table provides an overview of the treatments in the policy experiments and the number of participants in each treatment. In one round, we recruited 11,209 participants who were randomly allocated to state their level of agreement with government support for either males or females falling behind the labor market and education. Second, we elicited beliefs about the effort levels of those falling behind in the labor market and in education in two rounds (N=1,054 and N=1,001). In each round, the participants were randomly allocated to consider males or females falling behind.

Table S.3: Balance test, choice experiment, spectators

	Male	Republican	Income	Age
Mixed-gender merit male behind	-0.013	-0.010	1,846	0.079
	(0.012)	(0.011)	(1,397)	(0.385)
Mixed-gender luck female behind	0.001	-0.006	668	0.361
	(0.012)	(0.012)	(1,454)	(0.418)
Mixed-gender luck male behind	0.014	0.007	765	0.407
	(0.012)	(0.012)	(1,450)	(0.417)
Single-gender merit female behind	0.027	-0.014	530	0.536
	(0.036)	(0.033)	(4,279)	(0.968)
Single-gender merit male behind	0.037	0.002	1,864	1.319
	(0.036)	(0.034)	(4,705)	(0.931)
Single-gender luck female behind	0.009	0.026	2,216	-0.903
	(0.036)	(0.034)	(4,561)	(0.937)
Single-gender luck male behind	0.008	0.071**	-1,517	1.820*
	(0.036)	(0.035)	(4,402)	(0.931)
Included control:				
Round dummy	Yes	Yes	Yes	Yes
Observations	14,521	14,521	14,373	14,521
R^2	0.001	0.001	0.004	0.025
Prob > F	0.527	0.335	0.940	0.278

Note: The table reports linear regressions with the different spectator background characteristics as dependent variables. For each background characteristic, we have reported the p-value of the joint F-test testing whether the eight treatments are significantly different from each other with respect to that background characteristic. The reference category across all regressions is the mixed-gender merit treatment where a female has fallen behind. The dependent variable is an indicator variable for the spectator being male (column (1)), an indicator variable for the spectator being Republican (column (2)), the spectator's yearly household income (gross income before taxes are deducted, inflation adjusted to 2020) (column (3)), or the spectator's age in years (column (4)). The different income categories are specified in Appendix B.6. The spectators who did not know or want to state their income are not included in the sample in column (3). The regressions include indicator variables for the round of the study in which the spectator took part in (not reported). Robust standard errors in parentheses: * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

Table S.4: Balance test, choice experiment, effort beliefs

	Male	Republican	Income	Age
Mixed-gender merit	0.042	0.004	-4,815	0.138
male behind	(0.031)	(0.030)	(3,861)	(1.066)
Mixed-gender luck	-0.012	0.002	3,455	1.862
female behind	(0.031)	(0.030)	(4,210)	(1.367)
Mixed-gender luck	-0.006	-0.030	-265	2.650
male behind	(0.031)	(0.030)	(4,448)	(2.198)
Single-gender merit	0.024	0.022	-2,485	2.022
female behind	(0.031)	(0.030)	(3,981)	(1.844)
Single-gender merit	-0.008	0.036	-1,260	-1.592
male behind	(0.031)	(0.030)	(4,121)	(1.069)
Observations	3,000	3,000	3,000	3,000
R^2	0.002	0.002	0.001	0.003
Prob $>$ F	0.4567	0.3487	0.5010	0.0697

Note: The table reports linear regressions with the different participant background characteristics as dependent variables. For each background characteristic, we have reported the p-value of the joint F-test testing whether the six treatments are significantly different from each other with respect to that background characteristic. The reference category across all regressions is the mixed-gender merit treatment where a female has fallen behind. The dependent variable is "Male" (column (1)), "Republican" (column (2)), "Income" (column (3)), and "Age" (column (4)), defined in Table S.3. Robust standard errors in parentheses: * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

Table S.5: Balance test, policy experiment, government support

	Male	Republican	Income	Age
Male behind	0.010	0.010	2,572**	0.075
	(0.009)	(0.009)	(1,052)	(0.312)
Observations	11,209	11,209	10,838	11,209
R^2	0.000	0.000	0.001	0.000
Prob > F	0.2699	0.2681	0.0145	0.8094

Note: The table reports linear regressions with the different participant background characteristics as dependent variables for the sample in the policy experiment who are asked to what extent they agree with the government supporting people falling behind in the labor market and education. For each background characteristic, we have reported the p-value of the t-test testing whether the two treatments are significantly different from each other with respect to that background characteristic. The reference category across all regressions is the treatment in which the participants are asked about government support to females falling behind in the labor market and education. The dependent variable is "Male" (column (1)), "Republican" (column (2)), "Income" (column (3)), and "Age" (column (4)), defined in Table S.3. Robust standard errors in parentheses: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table S.6: Balance test, policy experiment, effort beliefs

	Male	Republican	Income	Age
Male behind	-0.048**	-0.013	2,573	-1.249
	(0.022)	(0.021)	(2,842)	(0.773)
Observations	2,055	2,055	2,055	2,055
R^2	0.002	0.000	0.000	0.001
Prob > F	0.029	0.534	0.365	0.106

Note: The table reports linear regressions with the different participant background characteristics as dependent variables for the sample in the policy experiment who are asked about the effort of people who fall behind in the labor market and education. For each background characteristic, we have reported the p-value of the t-test testing whether the two treatments are significantly different from each other with respect to that background characteristic. The reference category across all regressions is the treatment asking about females who have fallen behind. The dependent variable is "Male" (column (1)), "Republican" (column (2)), "Income" (column (3)), and "Age" (column (4)), defined in Table S.3. Robust standard errors in parentheses: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table S.7: Heterogeneity, choice experiment, spectator choice

Mixed-gender merit	Amount to worker behind (std)				
	Gender	Politics	Income	Age	All
Male behind	-0.185***	-0.144***	-0.117***	-0.151***	-0.179***
	(0.033)	(0.031)	(0.036)	(0.033)	(0.052)
Male participant	-0.025				-0.012
	(0.036)				(0.035)
Republican		-0.218***			-0.197***
		(0.037)			(0.037)
Low income			0.121***		0.097***
			(0.036)		(0.036)
Low age				0.115***	0.090**
				(0.036)	(0.036)
Male participant × Male behind	0.098*				0.093*
	(0.051)				(0.051)
Republican × Male behind		0.012			0.007
		(0.053)			(0.054)
Low income × Male behind			-0.042		-0.042
			(0.051)		(0.051)
Low age × Male behind				0.030	0.028
				(0.051)	(0.051)
Constant	0.094***	0.160***	0.022	0.027	0.067*
	(0.024)	(0.023)	(0.026)	(0.023)	(0.037)
Male behind + interaction	-0.087**	-0.132***	-0.159***	-0.121***	
	(0.039)	(0.043)	(0.036)	(0.038)	
Observations	7,243	7,243	7,243	7,243	7,243
R^2	0.006	0.016	0.008	0.010	0.021

Note: The table reports population-weighted linear regressions, where the dependent variable is the standardized amount transferred to the worker who is falling behind. The standardized variable is established on the basis of the choices made by the spectators in the mixed-gender merit treatments. "Male behind," "Male participant," "Republican," "Low income," and "Low age" are defined in Table 3. "Male participant × Male behind," "Republican × Male behind," "Low income × Male behind," and "Low age × Male behind" are interactions between the respective background characteristic variable and "Male behind". "Male behind+interaction" is a linear combination of "Male behind" and the respective interaction. The regressions include indicator variables for the round of the study in which the spectator took part in (not reported). Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S.8: Heterogeneity, choice experiment, effort beliefs

Mixed-gender merit	Level of agreement (std)				
	Gender	Politics	Income	Age	All
Male behind	0.400***	0.198**	0.283***	0.425***	0.597***
	(0.086)	(0.081)	(0.086)	(0.086)	(0.133)
Male participant	0.246**				0.211**
	(0.095)				(0.098)
Republican		0.004			0.013
		(0.100)			(0.101)
Low income			-0.086		-0.028
			(0.096)		(0.098)
Low age				0.447***	0.433***
				(0.094)	(0.093)
Male participant × Male behind	-0.323**				-0.339***
	(0.129)				(0.131)
Republican × Male behind		0.124			0.128
		(0.134)			(0.136)
Low income × Male behind			-0.070		-0.150
			(0.130)		(0.131)
Low age × Male behind				-0.389***	-0.333***
				(0.128)	(0.128)
Constant	-0.238***	-0.123**	-0.084	-0.331***	-0.416***
	(0.061)	(0.059)	(0.062)	(0.065)	(0.095)
Male behind + interaction	0.077	0.322***	0.213**	0.036	
	(0.096)	(0.107)	(0.097)	(0.095)	
Observations	1,000	1,000	1,000	1,000	1,000
R^2	0.023	0.017	0.019	0.040	0.053

Note: The table reports population-weighted linear regressions where the dependent variable is the standardized level of agreement with the worker falling behind having exerted less effort than the worker ahead. The standardized variable is established on the basis of the responses by the participants in the mixed-gender merit treatments. "Male behind," "Male participant," "Republican," "Low income," and "Low age" are defined in Table 3. "Male participant × Male behind," "Republican × Male behind," "Low income × Male behind," and "Low age × Male behind" are interactions between the respective background characteristic variable and "Male behind." "Male behind+interaction" is a linear combination of "Male behind" and the respective interaction. Robust standard errors in parentheses: * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

Table S.9: Choice experiment, spectator choice

Single-gender	Spectator choice				Effort beliefs			
	Nothing to worker behind		Amount to worker behind (std)		Agree low effort		Level of agreement (std)	
Male single-gender	-0.011	-0.014	-0.035	-0.029	0.008	0.006	0.054	0.050
	(0.042)	(0.042)	(0.082)	(0.081)	(0.033)	(0.032)	(0.066)	(0.064)
Luck	-0.112***	-0.115***	0.509***	0.515***				
	(0.040)	(0.040)	(0.088)	(0.087)				
Luck ×	0.005	0.002	0.135	0.142				
Male single-gender	(0.056)	(0.056)	(0.121)	(0.120)				
Male participant		0.039		-0.082		0.150***		0.234***
		(0.028)		(0.060)		(0.033)		(0.066)
Republican		0.098***		-0.219***		0.131***		0.281***
		(0.030)		(0.065)		(0.034)		(0.066)
Low income		0.010		-0.030		-0.102***		-0.225***
		(0.029)		(0.063)		(0.033)		(0.066)
Low age		-0.010		0.036		0.092***		0.160**
		(0.028)		(0.060)		(0.033)		(0.065)
Constant	0.332***	0.283***	-0.270***	-0.170**	0.492***	0.377***	-0.027	-0.207***
	(0.030)	(0.040)	(0.062)	(0.085)	(0.023)	(0.039)	(0.047)	(0.075)
Male behind (luck)	-0.006	-0.011	0.100	0.113				
	(0.037)	(0.037)	(0.088)	(0.088)				
Observations	1,026	1,026	1,026	1,026	999	999	999	999
R^2	0.015	0.029	0.084	0.098	0.000	0.058	0.001	0.052

Note: The table reports population-weighted linear regressions where the dependent variable is an indicator variable for the spectator transferring nothing to the worker who is falling behind (columns (1)–(2)), the standardized amount transferred to the worker who is falling behind (columns (3)–(4)), an indicator variable for the participant strongly or somewhat agreeing with the worker falling behind having exerted less effort than the worker ahead (columns (5)–(6)), and the standardized level of agreement with the worker falling behind having exerted less effort than the worker ahead (columns (7)–(8)). The standardized variables are established on the basis of the responses by the participants in the respective single-gender treatments. "Male single-gender" is an indicator variable for the spectator being in the single-gender treatment with two male workers and "Luck × Male single-gender" is an interaction between "Luck" and "Male single-gender." "Luck," "Male participant," "Republican," "Low income," and "Low age" are defined in Table 3. "Male behind (luck)" is a linear combination of "Male single-gender" and "Luck". The regressions include indicator variables for the round of the study in which the spectator took part in (not reported). Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S.10: Choice experiment, spectator choice

Single-gender vs.	Merit				Luck			
mixed-gender merit	Nothing to worker behind		Amount to worker behind (std)		Nothing to worker behind		Amount to worker behind (std)	
MG × Male behind	0.066**	0.066**	-0.126**	-0.124**	0.024	0.029	0.017	0.003
	(0.027)	(0.027)	(0.057)	(0.056)	(0.033)	(0.033)	(0.079)	(0.079)
MG × Female behind	-0.008	-0.008	0.055	0.055	0.025	0.026	0.040	0.039
	(0.027)	(0.027)	(0.057)	(0.057)	(0.033)	(0.033)	(0.078)	(0.078)
Constant	0.326***	0.303***	0.027	0.046	0.217***	0.176***	-0.014	0.148**
	(0.021)	(0.029)	(0.044)	(0.060)	(0.018)	(0.031)	(0.043)	(0.074)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,077	2,077	2,077	2,077	1,025	1,025	1,025	1,025
R^2	0.005	0.010	0.006	0.015	0.001	0.010	0.000	0.010

Note: The table reports population-weighted linear regressions where the dependent variable is an indicator variable for the spectator transferring nothing to the worker who is falling behind (columns (1)–(2) and (5)–(6)) and the standardized amount transferred to the worker who is falling behind (columns (3)–(4) and (7)–(8)). The standardized variables are established on the basis of the choices made by the spectators in the merit treatments (columns (3)–(4)) and the luck treatments (columns (7)–(8)) in the first and second round of the data collection. The reference category across all regressions is the single-gender merit treatments. "Male behind," "Male participant," "Republican," "Low income," and "Low age" are defined in Table 3. The four background characteristics are included as controls when specified. MG × Male behind is an indicator for the spectator being in a mixed-gender treatment where the male is falling behind and "MG × Female behind" is an indicator for the spectator being in a mixed-gender treatment where the female is falling behind. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table S.11: Heterogeneity, policy experiment, agreement with support

	Agreement with government support (std)				
	Gender	Politics	Income	Age	All
Male behind	-0.375***	-0.315***	-0.281***	-0.314***	-0.490***
	(0.025)	(0.021)	(0.025)	(0.025)	(0.036)
Male participant	-0.119***				-0.098***
	(0.027)				(0.026)
Republican		-0.634***			-0.595***
		(0.029)			(0.029)
Low income			0.135***		0.091***
			(0.027)		(0.026)
Low age				0.302***	0.232***
				(0.027)	(0.026)
Male participant × Male behind	0.212***				0.218***
	(0.038)				(0.036)
Republican × Male behind		0.163***			0.173***
		(0.041)			(0.040)
Low income × Male behind			0.025		0.043
			(0.038)		(0.037)
Low age × Male behind				0.093**	0.104***
				(0.037)	(0.036)
Constant	0.193***	0.326***	0.079***	-0.003	0.217***
	(0.018)	(0.015)	(0.018)	(0.018)	(0.026)
Male behind + interaction	-0.162***	-0.152***	-0.257***	-0.221***	
	(0.028)	(0.035)	(0.029)	(0.027)	
Observations	11,209	11,209	11,209	11,209	11,209
R^2	0.021	0.085	0.024	0.049	0.112

Note: The table reports population-weighted linear regressions where the dependent variable is the standardized level of agreement with the statement, "It is very important that the government provides support to males (females) who fall behind in education and in the labor market." The standardized variable is established on the basis of the responses by the participants who were presented with this statement. "Male behind" is an indicator variable for the participant being in the treatment describing males who fall behind. "Male participant," "Republican," "Low income" and "Low age" are defined in Table 3. "Male participant × Male behind," "Republican × Male behind," "Low income × Male behind" and "Low age × Male behind" are interactions between the respective characteristic and "Male behind". "Male behind+interaction" is a linear combination of "Male behind" and the respective interaction. Robust standard errors in parentheses, where * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

Table S.12: Heterogeneity, policy experiment, effort beliefs

	Agreement with lack of effort (std)				
	Gender	Politics	Income	Age	All
Male behind	0.472***	0.460***	0.448***	0.635***	0.636***
	(0.058)	(0.056)	(0.069)	(0.061)	(0.099)
Male participant	0.272***				0.188***
	(0.070)				(0.067)
Republican		0.313***			0.335***
		(0.074)			(0.070)
Low income			-0.289***		-0.216***
			(0.070)		(0.067)
Low age				0.504***	0.531***
				(0.069)	(0.067)
Male participant × Male behind	-0.004				0.017
	(0.090)				(0.088)
Republican × Male behind		0.026			-0.024
		(0.094)			(0.091)
Low income × Male behind			0.021		0.003
			(0.090)		(0.088)
Low age × Male behind				-0.330***	-0.315***
				(0.088)	(0.086)
Constant	-0.370***	-0.335***	-0.076	-0.480***	-0.587***
	(0.050)	(0.049)	(0.059)	(0.051)	(0.079)
Male behind + interaction	0.468***	0.486***	0.469***	0.305***	
	(0.069)	(0.076)	(0.058)	(0.064)	
Observations	2,054	2,054	2,054	2,054	2,054
R^2	0.071	0.077	0.073	0.088	0.142

Note: The table reports population-weighted linear regressions where the dependent variable is the standardized level of agreement with the statements; "When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort" or "Males (Females) falling behind in education and in the labor market have exerted low effort." The standardized variable is established on the basis of the responses by the participants who were presented with either of these statement. "Male behind" is an indicator variable for the participant being in the treatment describing males who fall behind. "Male participant," "Republican," "Low income," and "Low age" are defined in Table 3. "Male participant × Male behind," "Republican × Male behind," "Low income × Male behind" and "Low age × Male behind" are interactions between the respective characteristic and "Male behind." "Male behind+interaction" is a linear combination of "Male behind" and the respective interaction. Robust standard errors in parentheses, where * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

A.2 Multiple hypothesis testing

We here report the p-values adjusted for multiple hypothesis testing. We calculate unadjusted p-values as bootstrap p-values following Davison and Hinkley (1997) and compute p-values adjusted for stepdown multiple testing following the algorithm proposed by Romano and Wolf (2016). Bootstrapping is done with 10,000 replications.

Table S.13: Choice experiment, spectators: Average treatment effect (male behind vs. female behind), amount to worker behind (std)

	Treatment difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Mixed-gender			
Merit (male vs. female behind)	-0.130	0.000	0.000
Luck (male vs. female behind)	-0.064	0.020	0.065
Single-gender			
Merit (male vs. female behind)	-0.030	0.662	0.662
Luck (male vs. female behind)	0.102	0.216	0.379

Note: The multiple hypothesis adjustment is based on the following population-weighted linear regression specification

$$u_i = \alpha + \beta_1 T2_i + \beta_2 T3_i + \beta_3 T4_i + \beta_4 T5_i + \beta_5 T6_i + \beta_6 T7_i + \beta_7 T8_i + \gamma \mathbf{X}_i + \epsilon_i$$

where u_i is the standardized amount transferred by spectator i to the worker who is falling behind. $\mathbf{X}_i$ is a vector of control variables and includes indicator variables for gender, political party, income, and age, and round indicators as defined in Table 3; ϵ_i is an error term. The standardized variable is established on the basis of the choices made by the full set of spectators. Column (1) reports the estimated treatment effects (the difference in the standardized amount transferred to a male behind and a female behind). Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.14: Choice experiment, effort beliefs: Average treatment effect (male behind vs. female behind), agreement (std)

	Treatment difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Mixed-gender			
Merit (male vs. female behind)	-0.230	0.000	0.000
Luck (male vs. female behind)	-0.540	0.000	0.000
Single-gender			
Merit (male vs. female behind)	-0.044	0.534	0.534

Note: The multiple hypothesis adjustment is based on the following population-weighted linear regression specification

$$u_i = \alpha + \beta_1 T2_i + \beta_2 T3_i + \beta_3 T4_i + \beta_4 T5_i + \beta_5 T6_i + \gamma \mathbf{X}_i + \epsilon_i$$

where u_i is the dependent variable is the standardized level of agreement with the worker falling behind in the choice experiment having exerted low effort. $\mathbf{X}_i$ is a vector of control variables and includes indicator variables for gender, political party, income, and age as defined in Table 3; ϵ_i is an error term. The standardized variable is established on the basis of the responses by the participants who were asked about the effort level of the worker falling behind. Column (1) reports the estimated treatment effects (the difference in the standardized level of agreement with lack of effort being the reason why a male versus a female is behind). Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.15: Choice experiment, spectators: Average treatment effect in mixed-gender merit by subgroup, amount to worker behind (std)

	Treatment difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Female	-0.185	0.000	0.000
Male	-0.087	0.027	0.027
Republican	-0.132	0.003	0.005
Non-Republican	-0.144	0.000	0.000
High income	-0.117	0.002	0.004
Low income	-0.159	0.000	0.000
High age	-0.151	0.000	0.000
Low age	-0.121	0.002	0.005

Note: The multiple hypothesis adjustment is based on specification (2), one for each dimension: gender, political party, income, and age, and round indicators as defined in Table 3. The dependent variable is the standardized amount transferred to the worker who is falling behind. The standardized variable is established on the basis of the choices made by the spectators in the mixed-gender merit treatments. Column (1) reports the estimated treatment effects (the difference in the standardized amount transferred to a male behind and a female behind), by subgroup. Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.16: Choice experiment, spectators: Subgroup difference in average treatment effect in mixed-gender merit, amount to worker behind (std)

	Subgroup difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Males vs. females	0.098	0.056	0.199
Republicans vs. non-republicans	0.012	0.828	0.828
Low income vs. high income	-0.042	0.411	0.793
Low age vs high age	0.030	0.555	0.802

Note: The multiple hypothesis test is based on specification (2), including interaction variables and background variables for gender, political party, income and age, and round indicators as defined in Table 3. The dependent variable is the standardized amount transferred to the worker who is falling behind. The standardized variable is established on the basis of the choices made by the spectators in the mixed-gender merit treatments. Column (1) reports the estimated subgroup differences in treatment effect (the difference in the standardized amount transferred to a male behind and a female behind). Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.17: Choice experiment, effort beliefs. Average treatment effect in mixed-gender merit by subgroup, agreement (std)

	Treatment difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Female	0.400	0.000	0.000
Male	0.077	0.424	0.627
Republican	0.322	0.003	0.012
Non-Republican	0.198	0.015	0.047
High income	0.283	0.001	0.005
Low income	0.213	0.028	0.072
High age	0.425	0.000	0.000
Low age	0.036	0.708	0.708

Note: The multiple hypothesis adjustment is based on specification (2), one for each dimension: gender, political party, income, and age as defined in Table 3. The dependent variable is the standardized level of agreement with the worker falling behind having exerted less effort than the worker ahead. The standardized variable is established on the basis of the responses by the participants in the mixed-gender merit treatments. Column (1) reports the estimated differences in treatment effect (the difference in the standardized agreement when a male is behind and a female is behind), by subgroup. Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.18: Choice experiment, effort beliefs. Subgroup difference in average treatment effect in mixed-gender merit, agreement (std)

	Subgroup difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Males vs. females	-0.323	0.013	0.037
Republicans vs. non-republicans	0.124	0.355	0.582
Low income vs. high income	-0.070	0.589	0.589
Low age vs high age	-0.389	0.002	0.009

Note: The multiple hypothesis test is based on specification (2), including interaction variables and background variables for gender, political party, income and age as defined as defined in Table 3. The dependent variable is the standardized level of agreement with the worker falling behind having exerted less effort than the worker ahead. The standardized variable is established on the basis of the responses by the participants in the mixed-gender merit treatments. Column (1) reports the estimated subgroup differences in treatment effect (the difference in the standardized agreement when a male is behind and a female is behind). Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.19: Policy experiment, government support: Average treatment effect by subgroup, agreement (std)

	Treatment difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Female	-0.375	0.000	0.000
Male	-0.162	0.000	0.000
Republican	-0.152	0.000	0.000
Non-Republican	-0.315	0.000	0.000
High income	-0.281	0.000	0.000
Low income	-0.257	0.000	0.000
High age	-0.314	0.000	0.000
Low age	-0.221	0.000	0.000

Note: The multiple hypothesis adjustment is based on specification (2), one for each dimension: gender, political party, income, and age as defined in Table 3. The dependent variable is the standardized level of agreement with the statement, "It is very important that the government provides support to males (females) who fall behind in education and in the labor market." The standardized variable is established on the basis of the responses by the participants who were presented with this statement. Column (1) reports the estimated differences in treatment effect (the difference in the standardized level of agreement when males are behind and females are behind), by subgroup. Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.20: Policy experiment, government support. Subgroup difference in average treatment effect, agreement (std)

	Subgroup difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Males vs. females	0.212	0.000	0.000
Republicans vs. non-republicans	0.163	0.000	0.000
Low income vs. high income	0.025	0.512	0.512
Low age vs high age	0.093	0.015	0.029

Note: The multiple hypothesis test is based on specification (2), including interaction variables and background variables for gender, political party, income and age as defined in Table 3. The dependent variable is the standardized level of agreement with the statement, "It is very important that the government provides support to males (females) who fall behind in education and in the labor market." The standardized variable is established on the basis of the responses by the participants who were presented with this statement. Column (1) reports the estimated subgroup differences in treatment effect (the difference in the standardized level of agreement when males are behind and females are behind). Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.21: Policy experiment, effort beliefs: Average treatment effect by subgroup, agreement (std)

	Treatment difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Female	0.472	0.000	0.000
Male	0.468	0.000	0.000
Republican	0.486	0.000	0.000
Non-Republican	0.460	0.000	0.000
High income	0.447	0.000	0.000
Low income	0.469	0.000	0.000
High age	0.635	0.000	0.000
Low age	0.305	0.000	0.000

Note: The multiple hypothesis adjustment is based on specification (2), one for each dimension: gender, political party, income, and age, as defined in Table 3. The dependent variable is the standardized level of agreement with one of the statements, "When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort" or "Males (Females) falling behind in education and in the labor market have exerted low effort." The standardized variable is established on the basis of the responses by the participants who were presented with either of these statements. Column (1) reports the estimated treatment effects (the difference in the standardized level of agreement with lack of effort when males are behind and when females are behind), by subgroup. Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

Table S.22: Policy experiment, effort beliefs. Subgroup difference in average treatment effect, agreement (std)

	Subgroup difference	Unadjusted p-value	Romano-Wolf adjusted p-value
Males vs. females	-0.004	0.960	0.990
Republicans vs. non-republicans	0.026	0.788	0.990
Low income vs. high income	0.021	0.808	0.990
Low age vs high age	-0.330	0.000	0.001

Note: The multiple hypothesis test is based on specification (2), including interaction variables and background variables for gender, political party, income, and age as specified in Table 3. The dependent variable is the standardized level of agreement with one of the statements, "When males (females) fall behind in education and in the labor market, it largely reflects their lack of effort" or "Males (Females) falling behind in education and in the labor market have exerted low effort." The standardized variable is established on the basis of the responses by the participants who were presented with either of these statements. Column (1) reports the estimated subgroup differences in treatment effect (the difference in the standardized level of agreement with lack of effort when males are behind and when females are behind). Column (2) reports the unadjusted p-values and column (3) reports the Romano-Wolf adjusted p-values.

A.3 Pre-analysis plans

We implemented three rounds of the third-party spectator design, and registered pre-analysis plans at the AEA RCT Registry prior receiving data for each respective round. The largest data collection of spectator decisions was conducted in 2020, with 11,419 participants. We closely follow the pre-specified analysis for this data collection (AEARCTR-0005610), with some minor deviations:

1. We make semantic changes (changing labels) to make the paper and the results easier to read.

2. Because we do not find a significant difference in the gender bias between environments where luck is the source of inequality and where merit is the source of inequality in the full sample, we do not present how this varies by subgroup.

In addition to the data collection with 11,419 spectators, we collected data in two former rounds, with 2,052 and 1,051 spectators participating in each (AEARCTR-0000853 and AEARCTR-0001027). Since we base our analysis on the pre-specified empirical strategy adjusted to the largest data collection, we make some deviations from the two former pre-analysis plans:

1. We make semantic changes (changing labels, standardizing the dependent variable, changing from indicator for merit treatment to indicator for luck treatment) to make the paper and the results easier to read.

2. We make minor modelling changes to simplify the exposition of the results:

- As pre-specified for the main data collection, we base the main analysis on the full set of the mixed-gender treatments with interactions to identify separate treatment effects based on the source of inequality, instead of restricting samples by luck and merit treatments.
- We base the heterogeneity analysis on treatment×subgroup interactions instead of subgroup sample restrictions.
- The dependent variable was specified as the inequality implemented between the two workers in the first pre-analysis plan. This was changed to the standardized amount transferred to the worker behind, to capture the direction of any potential bias. We also flipped the alternative dependent variable from the share equalizing, to the share transferring nothing to the worker behind to simplify the exposition and interpretation of the results.

3. The single-gender treatments only amount to 6.9% of the final spectator sample, and we therefore do not present all the pre-specified analysis based on the single-gender treatments of the first data collection.

4. In the second round of data collection, we asked the participants about their support for affirmative action programs for women and their beliefs about the relative performance of 8th graders in mathematics and reading. Since these questions in the end only were asked to 7.2% of the spectator sample, we are not focusing on the analysis of these questions in the paper.

B Online Appendix: Sample and instructions

B.1 Data collections

The data for the choice experiment and the policy experiment was collected through three survey providers in a set of data collections. The survey providers, Kantar TNS, Ipsos and NORSTAT, are all international data collection agencies recruiting participants through large, existing panels with the aim of recruiting nationally representative samples on a set of observable characteristics (age, gender and broad geography). The variation in survey providers are a result of the legal requirement for competitive tenders as a public institution in Norway. The data was collected in six rounds.[1]

[1]We recruited the workers separately for each round of data collection on Mechanical Turk. In the first round of data collection, we recruited 1,370 workers, 685 men and 685 women living in the US. In the second round, we recruited 702 workers, 351 men and 351 women. In total, we ended up with 3,108 unique pairs of assignments/workers

Three rounds of third-party spectator design

The first round was the most comprehensive and included the full set of eight spectator treatments. The second round tested the robustness of the main treatment effects on the spectator choices in the mixed-gender merit treatments, beliefs about boys' and girls' relative performance in math and reading in school, and the spectators' support for affirmative action for females (the latter is not part of the main analysis, since the questions are only asked for a small share of the full sample). The results from the first two rounds of data collections in the choice experiment (3,102 participants) suggested that there is significant gender bias against males who are behind in mixed-gender environments where merit is the source of inequality ($p < 0.001$). The following data collection (11,419 participants) was designed and powered to study whether, in the mixed-gender environments, there is a significant difference in the gender bias between environments where luck is the source of inequality and where merit is the source of inequality. The goals of the third data collection, as described in the accompanying pre-analysis plan were:

- to check the robustness of the identified gender bias (in the first two rounds), and
- to test whether there is a significant interaction between the gender bias and the source of inequality (merit/luck) in the mixed-gender environment.

In the third data collection, we only collected data for Treatments 1–4, i.e. the mixed-gender treatments. The instructions for these four treatments are exactly the same as in the two former data collections with one exception. We now recruited 11,419 participants, and chose to implement 10% of the spectator decisions. The only difference between the instructions for the spectator decisions is therefore that we, in all treatments, add the following parenthesis at the end of the first sentence: "(one out of ten respondents to this survey are randomly selected and their choice will be implemented)."

For each round of data collection with third-party spectator decisions, we registered a pre-analysis plan (AEARCTR-0000853, AEARCTR-0001027 and AEARCTR-0005610, registered at the AEA RCT Registry). The main findings (Figure 1, Panel A) are robust across all rounds, and we control for the different rounds in the analysis.

Three rounds of effort beliefs and policy attitudes

In a fourth round of data collection, we also conducted an independent survey in which we elicited beliefs about effort, where each participant was randomly assigned to one of the treatments in

for the first two rounds of data collection. For the third round of data collection, we recruited 734 workers, 367 men and 367 women, which provided us with 1,101 unique pairs of assignments/workers, enough to implement about 10% of the spectator decisions in the third round of data collection.

the choice experiment. We recruited these participants separately from the spectators to avoid any confounds between the spectator decisions and elicited beliefs. We did not elicit beliefs for single-gender luck treatments, because in these treatments there is no basis for the spectator having different beliefs about the effort of the worker behind. In a fifth round of data collection, we elicited beliefs about the role of effort in explaining why people fall behind in education and in the labor market. To test the robustness of whether people are more likely to believe that males fall behind due to lack of effort in education and in the labor market, we elicited beliefs about the level of effort of people falling behind in a sixth round. However, the results hold for both questions separately and pooled. In a last round of data collection, we asked whether the participants agreed with the government providing support to people who fall behind in education and in the labor market. In all rounds and for all questions, we randomly manipulated the gender of those who fall behind.

An overview table is provided below:

Data collections				
Round	**Survey provider**	**Date**	**Sample size**	**# of treatments**
Choice experiment Spectators	Kantar TNS	Sept/Oct 2015	2,052	8
Choice experiment Spectators	Kantar TNS	Feb 2016	1,051	2
Choice experiment Spectators	Ipsos	Mar/Apr 2020	11,419	4
Choice experiment Beliefs	NORSTAT	Sept 2021	3,000	6
Policy experiment Beliefs I	Kantar TNS	Dec 2019	1,054	2
Policy experiment Beliefs II	NORSTAT	Sept 2021	1,001	2
Policy experiment Government support	NORSTAT	June/July 2022	11,209	2

B.2 Choice experiment, main spectator treatments

Treatment 1: Mixed-gender merit, female behind

In contrast to traditional survey questions that are about hypothetical situations, we now ask you to make a choice that has consequences for a real life situation. A few days ago two workers were recruited via an online labor market to conduct an assignment. They were both from the US; a man and a woman of the same age.

They were each paid a participation compensation of 2 USD regardless of what they would

end up being paid for the assignment. After completing the assignment, they were told that their earnings from the assignment would be determined by their productivity. The most productive worker would earn 6 USD for the assignment and the other worker would earn nothing for the assignment. They were not informed about who was the most productive worker. However, they were told that a third person would be informed about the assignment and who was the most productive worker. They were also told that this third person would be given the opportunity to redistribute the earnings and thus determine how much they were paid for the assignment.

You are the third person and we now want you to choose whether to redistribute the earnings for the assignment between the two workers. Your decision is completely anonymous. The workers will receive the payment that you choose for the assignment within a few days, but will not receive any further information.

The man was most productive and earned 6 USD for the assignment. The woman was least productive and earned nothing for the assignment.

Please state which of the following alternatives you choose:

I do not redistribute:

- *The most productive worker is paid 6 USD and the least productive worker is paid 0 USD.*

I do redistribute:

- *The most productive worker is paid X USD and the least productive worker is paid 6-X USD.* [where X is either 5, 4, 3, 2, 1 or 0]

B.3 Choice experiment, additional spectator treatments

Treatment 3: Mixed-gender, luck, female behind

Treatment 3 is identical to Treatment 1, except that we describe how a lottery determines earnings, instead of productivity, and therefore replace 'their productivity' with 'a lottery'; 'most productive worker' with 'worker winning the lottery'; 'who was the most productive worker' with 'the outcome of the lottery; 'lucky' with 'most productive', and 'unlucky' with 'least productive.' To describe a female behind instead of a male behind, we also replace the gender (man/woman) in the description of who was lucky and unlucky directly before the choice alternatives.

Treatment 5: Single-gender, merit, two females

Identical to Treatment 1, except that we describe the two workers as two women instead of a man and a woman. In identifying who was the most productive, the instructions are the following: *One*

of the women was most productive and earned 6 USD for the assignment. The other woman was least productive and earned nothing for the assignment.

B.4 Choice experiment, beliefs and attitudes questions

Choice experiment, beliefs about effort

We asked the following questions to a general population sample of 3,000 Americans, recruited through the data collection agency NORSTAT in September 2021. Each participant was randomly assigned to one of six treatments, where we varied the gender composition of the workers, the gender of the person falling behind, and the information provided on the source of inequality. Together, treatments 1–6 describe the environments in the mixed-gender merit treatments, the mixed-gender luck treatments, and the single-gender merit treatments, respectively.

The response scale for each question: Strongly disagree/ Somewhat disagree/ Neither agree nor disagree/ Somewhat agree/ Strongly agree.

Treatment 1 (male behind): In a recent study, we recruited two workers via an online labor market to conduct an assignment. They were both from the US; a man and a woman of the same age. Before they did the assignment, they were told that they would be paid a participation compensation of 2 USD regardless of what they would end up being paid for the assignment. However, they were not informed about how their earnings from the assignment would be determined. The man was less productive than the woman on the assignment.

We would like to know the extent to which you agree with the following statement:

"I expect that the less productive man exerted less effort on the assignment than the more productive woman."

Treatment 2 (female behind): Identical to Treatment 1, except that 'man' and 'woman' replace each other in the third last- and last sentence.

Treatment 3 (male behind): Identical to Treatment 1, except that the following sentence is removed: "The man was less productive than the woman on the assignment," and the statement the participants are to evaluate no longer mentions productivity: "I expect that the man exerted less effort than the woman on the assignment."

Treatment 4 (female behind): Identical question as in Treatment 3, except that 'man' is replaced with 'woman' in the last sentence.

Treatment 5 (male behind): Identical to Treatment 1, except that there are 'two men' working instead of 'a man and a woman,' the following sentence is removed: "The man was less productive than the woman on the assignment," and 'woman' is replaced by 'man' in the last sentence.

Treatment 6 (female behind): Identical question as in Treatment 5, except that 'man' and 'men' are switched out with 'woman' and 'women,' respectively.

Choice experiment, spectator data collection 2

Question 1) US 8th graders were tested in mathematics. How do you think the male students performed relative to the female students in a) mathematics? (Males much better/ Males somewhat better/ Equal performance/ Females somewhat better/ Females much better)

We then asked the same question about reading (1b), by inserting reading instead of mathematics in the question structure above.

Question 2) Do you generally favor or oppose affirmative action programs for women? (Generally favor/ Generally oppose)[2]

B.5 Policy experiment

Policy experiment, effort explains

We asked the following questions to a general population sample of 1,054 Americans, recruited through the data collection agency Kantar TNS in December, 2019. Each participant was randomly assigned one of the questions.

The response scale for each question: Strongly disagree/ Somewhat disagree/ Neither agree nor disagree/ Somewhat agree/ Strongly agree.

Treatment 1 (males behind): In the US, some males fall behind in education and in the labor market. To what extent do you agree with the statement: "When males fall behind in education and in the labor market, it largely reflects their lack of effort."

Treatment 2 (females behind): Identical question as in Treatment 1, except that 'males' is replaced with 'females'.

Policy experiment, level of effort

We asked the following questions to a general population sample of 1,001 Americans, recruited through the data collection agency NORSTAT in September 2021. Each participant was randomly assigned one of the questions.

The response scale for each question: Strongly disagree/ Somewhat disagree/ Neither agree nor disagree/ Somewhat agree/ Strongly agree.

[2]Question 1a and 1b are based on a nationally representative assessment conducted in 2015 by The National Assessment of Educational Progress (NAEP). Here, male and female 8th graders performed equally well on the mathematics test (each group with an average of 282 points), while females on average performed slightly better than males on the reading test (271 vs. 260 points). For more details, see http://www.nationsreportcard.gov/reading_math_2015/. Question 2 is taken from Gallup's Minority Rights and Relations survey conducted in 2015 with more than 2000 US adults. They found that 67% of Americans are in favor of affirmative action for women, with females being more prone to support it than males (72% vs. 62%).

Treatment 1 (males behind): We would like to know the extent to which you agree with the following statement: "Females falling behind in education and in the labor market have exerted low effort."

Treatment 2 (females behind): Identical question as in Treatment 1, except that 'Males' is replaced with 'Females'.

Policy experiment, government support

We asked the following questions to a general population sample of 11,209 Americans, recruited through the data collection agency NORSTAT in June–July 2022. Each participant was randomly assigned one of the questions.

The response scale for each question: Strongly disagree/ Somewhat disagree/ Neither agree nor disagree/ Somewhat agree/ Strongly agree.

Treatment 1 (males behind): In the US, some males fall behind in education and in the labor market. We would like to know the extent to which you agree with the following statement: "It is very important that the government provides support to males who fall behind in education and in the labor market."

Treatment 2 (females behind): Identical question as in Treatment 1, except that 'males' is replaced with 'females'

B.6 Background questions

In addition, the participants answered the following set of background questions:

Choice experiment, spectator data collection 1 and 2

- (...) we would be grateful if you could type in your actual age below?
- Are you? (Male/Female)
- In which region do you live? (Northeast/ Midwest/ South/ West)
- What is your household's combined yearly income (gross income before taxes are deducted)? (9 intervals listed from 'Less than $20,000' to '$150,000 or more'/ Do not know or prefer not to state)
- Which political party would you vote for if there was an election tomorrow? (Republican/ Democratic/ Other)

Choice experiment, spectator data collection 3

- What is your date of birth?[3]
- What is your gender?
- Please insert your zip code:[4]
- What is the highest degree or level of school you have completed? (Education through Grade 12 (Grade 4 or less/ Grade 5 to 8/ Grade 9 to 11/ Grade 12 (no diploma))/ High School Graduate (Regular High School Diploma/GED or alternative credential)/ College or Some College (Some college credit, but less than 1 year/ 1 or more years of college credit, no degree/ Associate's degree (AA, AS, etc.)/Bachelor's degree (BA, BS, etc.)/ After Bachelor's Degree (Master's Degree (MA, MS, MBA, etc.) or Professional degree (MD, DDS, JD, etc.) or Doctorate degree (PhD, EdD, etc.))
- Please indicate your annual household income before taxes. (23 intervals listed from 'Less than $5,000' to '$250,000 or more'/ Prefer not to answer)
- Which political party would you vote for if there was an election tomorrow? (Republican/ Democratic/ Other/ Prefer not to answer)

Policy experiment, beliefs about effort, round 1

- (...) we would be grateful if you could type in your actual age below?
- Are you? (Male/ Female)
- In which region do you live? (State in the United States)
- Please indicate your annual household income before taxes. (15 intervals listed from 'Less than $10,000' to '$200,000 or more'/ Don't know or prefer not to state)
- Which political party would you vote for if there was an election tomorrow? (Republican/ Democratic/ Other/ Prefer not to answer)

Policy experiment, beliefs about effort, round 2 and Choice experiment, effort beliefs

- Please enter your age
- Are you? (Male/ Female)
- Which state do you live in? (Dropdown menu of US states)

[3]We only have access to age in years.
[4]We only have access to state-level information.

- What is the highest degree or level of school you have completed? (Education through Grade 12 (Grade 4 or less/ Grade 5 to 8/ Grade 9 to 11/ Grade 12 (no diploma))/ High School Graduate (Regular High School Diploma/GED or alternative credential)/ Some college, no degree/ Associate's degree (AA, AS, etc.)/ Bachelor's degree (BA, BS, etc.)/ After Bachelor's Degree (Master's Degree (MA, MS, MBA, etc.)/ Professional degree (MD, DDS, JD, etc.)/ Doctorate degree (PhD, EdD, etc.))
- Please indicate your annual household income before taxes. (25 intervals listed from 'Less than $5,000' to '$250,000 or more')
- Which political party would you vote for if there was an election tomorrow? (Republican/ Democratic/ Other/ Prefer not to answer)
- Is religion an important part of your life? (Yes/ No/ Do not wish to respond)

Policy experiment, beliefs, round 2 and Choice experiment, beliefs

- Please indicate your gender. (Male/ Female)
- Please indicate your age.
- Where do you live? Select state. (Dropdown menu of US states)
- Please state your annual household income: (13 intervals listed from 'Less than $20,000' to 'Over $200,000')/ Would rather not say)
- What is your highest completed level of education? (Completed some high school/ High school graduate or GED equivalent/ Completed some college/ Associates degree/ College degree/ Completed some postgraduate/ Master's degree/ Doctorate degree/ None of the above)
- If there was a presidential election tomorrow, which party would you vote for? (The Republican Party/ The Democratic Party/ An independent party/ Do not want to answer/ Do not know/ Not eligible to vote)

References

Davison, A.C. and D. V. Hinkley (1997). *Bootstrap methods and their application*, Cambridge University Press.

Romano, Joseph P. and Michael Wolf (2016). "Efficient computation of adjusted p-values for resampling-based stepdown multiple testing," *Statistics & Probability Letters*, 113(1): 38–40.

Issued in the series Discussion Papers 2022

2022

01/22 January, **Astrid Kunze**. "Parental leave and maternal employment"

02/22 February, Øystein Foros, Hans Jarle Kind and Frank Stähler. "Endogenous multihoming and network effects: Playstation, Xbox, or both?"

03/22 March, **Catalina Franco**, David J. Harding, Shawn D. Bushway and Jeffrey Morenoff. "Failing to Follow the Rules: Can Imprisonment Lead to More Imprisonment Without More Actual Crime?"

04/22 March, Anselm Hager and **Justin Valasek**. "The Impact of Exposure to Refugees on Prosocial Behavior"

05/22 March, Anselm Hager and **Justin Valasek**. "The Impact of Forced Migration on In-Group and Out-Group Social Capital"

06/22 March, **Kurt R. Brekke**, Dag Morten Dalen and Odd Rune Straume. "Competing with precision: incentives for developing predictive biomarker tests"

07/22 May, **Alexander W. Cappelen**, Thomas de Haan and **Bertil Tungodden**. "Fairness and limited information: Are people Bayesian meritocrats?"

08/22 June, Rita Ginja, Julie Riise, Barton Willage and **Alexander Willén**. "Does Your Doctor Matter? Doctor Quality and Patient Outcomes"

09/22 June, **Richard Friberg**, **Emil M. Strøm Halseth**, **Frode Steen** and **Simen A. Ulsaker.** "The effect of cross-border shopping on commodity tax revenue: Results from a natural experiment"

10/22 July, **Alexander W. Cappelen**, **Ranveig Falch**, **Zhongjing Huang** and **Bertil Tungodden**. "Acceptance of inequality between children: Large-Scale Experimental Evidence from China and Norway"

11/22 August, **Øivind A. Nilsen** and Arvid Raknerud. "Dynamics of First-Time Patenting Firms"

12/22 September, Pedro Carneiro, **Kjell G. Salvanes**, Barton Willage, and **Alexander Willén**. "The Timing of Parental Job Displacement, Child Development and Family Adjustment"

13/22 October, Gozde Corekcioglu, Marco Francesconi and **Astrid Kunze**. "Expansions in Paid Parental Leave and Mothers' Economic Progress"

14/22 October, Ryan McWay, **Pallavi Prabhakar** and Ayo Ellis. "The Impact of Early Childhood Development Interventions on Children's Health in Developing Countries: A Systematic Review and Meta-analysis"

15/22 October, Lars Ivar Oppedal Berge, **Kjetil Bjorvatn**, Fortunata Makene, Linda Helgesson Sekei, **Vincent Somville** and **Bertil Tungodden**. "On the Doorstep of Adulthood: Empowering Economic and Fertility Choices of Young Women"

16/22 October, **Astrid Kunze** and Katrin Scharfenkamp. "Gender diversity, labour in the boardroom and gender quotas"

17/22 November, **Samuel Dodini**, **Katrine Løken**, and **Alexander Willén**. "The Effect of Labor Market Competition on Firms, Workers, and Communities"

18/22 November, **Yuzuru Kumon** and Kazuho Sakai. "Women's Wages and Empowerment: Pre-industrial Japan, 1600-1890"

19/22 December, **Alexander Willén**, Barton Willage, and Julie Riise. "Employment Protection and Child Development"

20/22 December, **Samuel D. Hirshman** and **Alexander Willén**. "Does Increasing Risk Widen Gender Gaps?"

21/22 December, Manudeep Bhuller, Gordon B. Dahl, **Katrine V. Løken** and Magne Mogstad. "Domestic Violence and the Mental Health and Well-being of Victims and Their Children"

2023

01/23 January, Felix Chopra, **Ingar Haaland** and Christopher Roth. "The Demand for News: Accuracy Concerns versus Belief Confirmation Motives"

02/23 March, **Ola Honningdal Grytten** and Viktoriia Koilo. "Offshore Industries as Growth Sector: The Norwegian Case"

03/23 March, **Ingvild Almås**, **Alexander W. Cappelen**, **Erik Ø. Sørensen** and **Bertil Tungodden**. "Fairness Across the World"

04/23 March, **Kurt R. Brekke**, Dag Morten Dalen and Odd Rune Straume. "The price of cost-effectiveness thresholds"

05/23 March, Leonardo Bursztyn, **Alexander W. Cappelen**, **Bertil Tungodden**, Alessandra Voena and David Yanagizawa-Drott. "How Are Gender Norms Perceived?"

06/23 April, **Alexander W. Cappelen**, Stefan Meissner and **Bertil Tungodden**. "Cancel the deal? An experimental study on the exploitation of irrational consumers"

07/23 May, **Sara Abrahamsson**, **Aline Bütikofer** and Krzysztof Karbownik. "Swallow this: Childhood and Adolescent Exposure to Fast Food Restaurants, BMI, and Cognitive Ability"

08/23 May, Malin Arve and **Justin Valasek**. "Underrepresentation, Quotas and Quality: A dynamic argument for reform"

09/23 May, **Björn Bartling**, **Alexander W. Cappelen**, Henning Hermes, Marit Skivenes and **Bertil Tungodden**. "Free to Fail? Paternalistic Preferences in the United States"

10/23 May, Chiara Canta, **Øivind A. Nilsen** and **Simen A. Ulsaker**. "Competition and risk taking in local bank markets: evidence from the business loans segment"

11/23 May, **Aline Bütikofer**, **René Karadakic** and **Alexander Willén**. "Parenthood and the Gender Gap in Commuting"

12/23 May, **Samuel Dodini**, **Kjell G. Salvanes**, **Alexander Willén** and **Li Zhu**. "The Career Effects of Union Membership"

13/23 June, **Alexander W. Cappelen**, **Ranveig Falch** and **Bertil Tungodden.** "Experimental Evidence on the Acceptance of Males Falling Behind"

NORGES HANDELSHØYSKOLE
Norwegian School of Economics

Helleveien 30
NO-5045 Bergen
Norway

T +47 55 95 90 00
E nhh.postmottak@nhh.no
W www.nhh.no