THE EFFECT OF UNFAIR CHANCES AND GENDER DISCRIMINATION ON LABOR SUPPLY*

Nickolas Gagnon Kristof Bosmans Arno Riedl

August 2020

Abstract

Labor market opportunities and wages may be unfair for various reasons, and how workers respond to different types of unfairness can have major economic consequences. Using an online labor platform, where workers engage in an individual task for a piece-rate wage, we investigate the causal effect of neutral and gender-discriminatory unfair chances on labor supply. We randomize workers into treatments where we control relative pay and chances to receive a low or a high wage. Chances can be fair, unfair based on an unspecified source, or unfair based on gender discrimination. Unequal pay reduces labor supply of low-wage workers, irrespective of whether the low wage is the result of fair or unfair chances. Importantly, the source of unfair chances matters. When a low wage is the result of gender-discriminatory chances, workers matched with a high-wage worker substantially reduce their labor supply compared to the case of equal low wages (-22%). This decrease is twice as large as those induced by low wages due to fair chances or unfair chances coming from an unspecified source. An additional experiment confirms the deleterious effect of gender discrimination on labor supply in a work environment devoid of chances, and highlights that workers' beliefs about facing discrimination matter for their responses. Our results concerning gender discrimination indicate a new reason for the lower labor supply of women, which is a prominent explanation for the gender gap in earnings. (*JEL*: D90, E24, J22, J31, J71, M5)

Keywords: Labor Supply; Wage Inequality; Procedural Fairness; Gender Discrimination

*We are grateful to audiences at seminars in Berlin, Heidelberg, Lyon, Maastricht, Middlebury, Montréal (CIRANO, HEC, and Concordia), NYU Abu Dhabi, Québec, Rotterdam, Toronto and Tucson, and at the ASFEE (Toulouse), Conference on Decision Sciences (Konstanz), ESA (Los Angeles), HDPD Retreat (Heerlen), IMEBEES (Utrecht), KVS (Leiden), M-BEPS (Maastricht), Paris2-ZEW Workshop on Market Design (Mannheim), and SABE (Moscow) for very useful comments. We are also thankful to Charles Bellemare, Pol Campos i Mercadé, Martin Dufwenberg, Bernard Fortin, Gautam Gowrisankaran, Olivier Marie, Pierre-Carl Michaud, Leticia Micheli, Rebecca Morton, Marcello Negrini, Charles Noussair, Juan Palacios, Noémi Péter, Gerhard Riener, Christina Rott, Arthur Schram, Peter Werner, and Henrik Zaunbrecher for their valuable comments. Nickolas Gagnon acknowledges financial support from the Fonds de recherche du Québec - Société et culture, and from the Human Decisions & Policy Design Research Theme of Maastricht University. Authors Information: Department of Microeconomics and Public Economics, School of Business and Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Nickolas Gagnon (corresponding author): n.gagnon@maastrichtuniversity.nl. Kristof Bosmans: k.bosmans@maastrichtuniversity.nl. Arno Riedl: a.riedl@maastrichtuniversity.nl. We obtained ethical approval from the Ethical Review Committee Inner City of Maastricht University (ID: ERCIC-054-15-11-2017), and pre-registered our hypotheses in the American Economic Association's Randomized Control Trials Registry (ID: AEARCTR-0002655 and AEARCTR-0003379).

I. Introduction

Chances are a pervasive feature of labor market activities and outcomes. Workers face them in hiring and promotion processes, when being up for bonus payments, and when being confronted with dismissal decisions. Such chances might be fair, but can also be unfair for many reasons, a prominent one being gender discrimination.[1] Accordingly, gender discrimination ranks high on economists' research agenda. However, the bulk of the research on the topic focuses on the demand side, leaving potential supply-side effects underexplored. Using experiments on an online labor platform, we present the first causal evidence on the impact of wage-related unfair chances and gender discrimination on labor supply decisions of workers.

Recent work has shown that unequal pay can have major effects on workers' behavior, such as reduced labor supply and productivity as well as increased job separations (Bracha et al., 2015; Breza et al., 2018; Dube et al., 2019). At the same time, a lasting idea in economics is that unfair chances influence equity judgments (e.g., Diamond, 1967), and empirical studies have provided support for this claim (e.g., Bolton et al., 2005; Brock et al., 2013; Cappelen et al., 2013). Together, these two strands of literature suggest that unequal chances in labor markets may also affect workers' behavior. Moreover, unequal chances and wages may generate especially strong responses if they are due to gender discrimination.

To investigate the causal effect of unfair chances and gender discrimination on labor supply, we conduct two experimental studies. All our experiments use an online labor platform, where we hire workers who individually engage in a task at a fixed piece-rate wage. The two studies differ in their main research objectives. In the main study (Study 1) we can draw causal conclusions on the effect of equal versus unequal chances and on the effect of the source of unequal chances, which can be neutral or explicitly gender discriminatory. In the follow-up study (Study 2), we investigate the labor supply effects of wage inequality due to implicit and explicit gender discrimination as well as the effect of beliefs about the presence of gender discrimination. In what follows we first describe the main features and findings of Study 1.

In the experiment, workers first learn the chances that will determine their and another worker's wage, which can be high or low. After being informed of their own and the other worker's realized wage, they decide individually how much to work. To cleanly isolate the effect of chances and their source on labor supply, we implement a design that rules out peer interactions, risk, and reciprocity considerations. Providing less labor is costly to workers because it reduces their own earnings, but has no other effects. Moreover, the use of an online labor platform provides us with a relatively large number of participants who are arguably

[1] See the review by Blau and Kahn (2017) on the gender gap in earnings, including the gap in promotion chances. Recent evidence also shows that women evaluated for tenure in economics departments face lower chances than men of equal ability (Sarsons et al., 2019). More generally, 80% of UK female employees believe that gender discrimination exists in the workplace (Investors in People, 2018) and 42% of their US counterparts have faced such workplace discrimination (Parker and Funk, 2017). For recent reviews of discrimination experiments, see, for instance, Lane (2016), Bertrand and Duflo (2017), and Neumark (2018).

more representative than student participants. The anonymity of the online platform also helps in avoiding peer effects from observing other workers quitting the task.[2]

We measure labor supply in the real effort task and implement five different payment schemes (treatments) where we create wage inequality through fair or unfair chances and also vary the source of unfair chances. We randomize each worker into one of the five schemes. In each scheme, a worker is anonymously paired with another worker engaged in the same individual task. In two baseline schemes, there is no wage inequality and both workers receive either the low wage or the high wage. In these schemes, we do not mention any chances that determine the wages. In the other three schemes workers in a pair receive unequal wages: one receives the high wage and one receives the low wage. In one payment scheme, both workers have a fair chance (50%) of receiving the high wage and the source of their chances is not mentioned. In the two other schemes one worker has a 25% chance of receiving the high wage and the other has a 75% chance. These two schemes differ in the information on the source of the unfair chances. In one scheme no specific source is mentioned. In the other scheme pairs of workers consist of a man and a woman, who are informed that their chances depend on their gender. We implement an explicit gender-discriminatory policy, to fix beliefs about gender discrimination.

To derive hypotheses, we provide a theoretical framework that extends the model of inequality aversion and work morale by Card et al. (2012) and Breza et al. (2018) with the model of social preferences incorporating chances by Saito (2013).[3] In our framework, both unfair chances and wage inequality increase the marginal disutility of working, thereby reducing labor supply. In addition, we posit that gender discrimination as the source of unfair chances creates an extra psychological cost that further increases workers' marginal disutility from unequal treatment.

We compare the labor supply of workers at a given wage across payment schemes, and test four pre-registered hypotheses. Our first and second hypotheses state that for low-wage workers and high-wage workers, respectively, unequal wages, unfair chances, and gender discrimination each incrementally decrease labor supply. Our third hypothesis says that these effects are stronger for low-wage workers than for high-wage workers. Finally, our fourth hypothesis conjectures that men and women may not react equally to gender-discriminatory chances compared to the same unfair chances emanating from an unspecified source.

Our main results can be summarized as follows. The first hypothesis is partly supported. Specifically, as expected, low-wage workers who are matched with a high-wage worker through fair chances supply significantly less labor (−13%) than low-wage workers who are matched with another low-wage worker. This effect of wage inequality is consistent with results reported in the three previous studies investigating the effect of wage inequality without chances and discrimination (Bracha et al., 2015; Breza et al., 2018; Dube et al., 2019). Our result adds that the unequal wage effect does not disappear when introducing a fair chance to receive the low or high wage. Further, we find that low wages resulting from unfair chances based on an

[2]In Section IV.B we discuss potential drawbacks of online experiments and how we minimize them.

[3]See also, Trautmann (2009).

unspecified source do not have an additional effect on labor supply, which partly contradicts our first hypothesis. However, workers do significantly decrease their labor supply if their low wage is due to unfair chances based on gender discrimination. The average impact of gender discrimination on labor supply is significant and large in economic terms. Low wages resulting from unfair gender-discriminatory chances reduce labor supply by 15% relative to low wages coming from unfair chances based on an unspecified source, and by 22% relative to equal low wages.

Interestingly, high-wage workers appear to be immune to the different types of inequality that we employ. Their labor supply does not significantly differ across payment schemes. Thus, the second hypothesis is not supported.

The third hypothesis, that the adverse effects on labor supply are stronger for low-wage workers than for high-wage workers, is supported when comparing the payment scheme with gender-discriminatory chances to the other schemes. In all other comparisons, low-wage workers and high-wage workers do not significantly differ in their responses.

Finally, we find that low-wage women respond more strongly to gender-discriminatory unfair chances than do low-wage men, supporting our fourth hypothesis. Moreover, an exploratory comparison across all five payment schemes reveals a distinct gender difference among low-wage workers. Men decrease their labor supply in response to any type of disadvantageous inequality, whereas disadvantaged women reduce their labor supply only if the low-wage is due to gender-discriminatory chances.

Our Study 2 builds upon and extends our main study. First, the second study aims to explore more cleanly the potentially different responses of men and women by disregarding the effects of chances. Second, in Study 1 gender discrimination was made explicit in order to control for beliefs. However, arguably, gender discriminatory practices in the field are more subtle and implicit, giving beliefs about the use of gender discrimination some room to affect labor supply. A second aim of Study 2 is to explore the difference between implicit and explicit gender discrimination and the role played by beliefs. Hypotheses were also pre-registered for this study.

The second study differs from the main study in the following aspects. First, unequal wages in a worker pair were implemented directly without reference to chances. Second, there were three different payment schemes. In one scheme, workers in a pair were informed about their own and the other worker's wage without any reference to gender. In a second scheme, workers were informed about the wages and about the gender of the other worker, who was always of opposite gender. This constitutes an *implicit* gender discrimination payment scheme. In a third scheme, workers were informed that they receive the high or low wage because of their gender. This constitutes an *explicit* gender discrimination payment scheme. Third, in each payment scheme, at the end of the experiment we asked workers how much they believed that their wage was based on gender discrimination.

The main results for Study 2 can be briefly summarized as follows. First, we corroborate the result that, on average, explicit negative gender discrimination decreases labor supply, but we do not find an effect of implicit gender discrimination. Second, this result also holds when looking at both genders separately. Third, we find limited evidence for that positive discrimination increases labor supply of women; implicit discrimination (marginally) increases their labor supply, but explicit discrimination has no effect. In contrast, positive discrimination of men decreases their labor supply. Fourth, there is a negative correlation between labor supply and the belief that gender discrimination was used to determine wages in the experiment, especially for women. This strongly suggests that the belief about the presence of gender discrimination is an important factor when the use of discrimination is not plainly obvious.

Our paper provides a number of contributions to the literature. First, it is the first to investigate the causal effect of unfair chances on labor supply decisions of workers. Our finding that the fairness of chances from an unspecified source does not affect labor supply stands in contrast to the empirical literature on unfair chances and income redistribution (e.g., Bolton et al., 2005; Krawczyk and Le Lec, 2010; Brock et al., 2013; Cappelen et al., 2013). Specifically, whereas the distribution of initial chances has been shown to influence redistribution decisions, we find that it has no impact on workers' labor supply decisions. This suggests that the response to unfair chances does not carry over automatically across decision domains.[4]

Second, our paper also provides the first evidence of the causal impact of gender discrimination on labor supply decisions. Our finding that unfair chances based on gender and negative gender-discriminatory wages both have a large detrimental impact on labor supply adds to the few economic studies showing that (ethnic minority) workers modify their behavior when discriminated (Parsons et al., 2011; Glover et al., 2017). Our study is the only one that identifies the effect of discrimination while controlling for changes in monetary incentives resulting from the presence of discrimination.

A third contribution of our paper is to the research on gender differences in labor markets more generally (Croson and Gneezy, 2009; Niederle, 2016). Our result that men reduce their labor supply in reaction to unequal wages from fair and unfair chances stemming from an unspecified source, whereas women do not, is consistent with the findings of Bracha et al. (2015). They, however, do not study the effect of unfair chances. In addition, we show that women also decrease their labor supply if they face lower wages that are the result of negative gender discrimination. We also provide the first tentative evidence that implicit positive discrimination of women can increase their labor supply, although explicit discrimination does not (and positive discrimination of men always decreases their labor supply). Moreover, we show that beliefs about the presence of negative discrimination may be pivotal in affecting labor supply when discrimination is not obvious.

Finally, an important implication of our paper is that it suggests a novel and complementary explanation for the gender gap in earnings. Goldin (2014) and Blau and Kahn (2017) review the literature on the earnings gender gap and conclude that the modern gender gap is mostly explained by the lower labor supply of women. The standard explanation for this lower labor

[4]The random procedure that we use to assign chances from an unspecified source is identical to the one used in several of the cited papers. Therefore, it is unlikely to explain why workers do not respond to unfair chances.

supply is that women value temporal flexibility more than men, plausibly because they have to bear greater household responsibilities (e.g., Goldin, 2014; Wiswall and Zafar, 2017). We provide evidence that the experience of (or belief in the existence of) gender discrimination itself can reduce labor supply. Not accounting for this channel might lead to misjudgment of the impact of discrimination and ill-advised policies. The identified channel may also affect the gender earnings gap in the manner of a self-fulfilling prophecy. Since the labor market offers high returns to long work hours, a lower labor supply can be both a reaction to (believed) discrimination and a rationale for employers to pay women less than men.

The rest of the paper is organized as follows. Section II situates our research in the literature, Section III presents the theoretical framework, Section IV describes the design of our main study, Section V provides the hypotheses, Section VI reports the results, Section VII describes and reports on the second study, Section VIII discusses our findings and their implications, and Section IX briefly concludes.

II. Related Literature

II.A. Unequal Wages

A stream of literature suggests that wage differentials perceived as unfair hamper the work morale (Adams, 1965; Akerlof and Yellen, 1990; Pfeffer and Langton, 1993; Bewley, 1999). Empirical evidence shows that unequal wages for similar work indeed negatively affect several labor outcomes. For instance, wage inequality decreases work satisfaction and increases job searches among disadvantaged workers (Card et al., 2012) and hurts their productivity (Gächter and Thöni, 2010; Cohn et al., 2014; Ockenfels et al., 2014).[5]

Three recent studies investigate the effect of unequal wages on labor supply, which is our variable of interest. Bracha et al. (2015) report a laboratory experiment where workers are paid piece-rate wages in an individual task and have to decide for how long to work. If no justification is provided, unequal wages decrease labor supply of male but not of female low-pay workers. High-pay workers' labor supply does not respond to wage inequality, irrespective of gender. Breza et al. (2018) conduct a field experiment with male workers in an Indian firm. Workers work individually in small teams in which wage inequality is manipulated. The authors find that wage inequality has no effect when productivity differences are observable. In contrast, when productivity differences are unobservable, then inequality decreases labor supply of low-pay and high-pay workers. Dube et al. (2019) exploit a natural experiment caused by changes in the wage structure of a large American firm. They find that workers arbitrarily receiving a low relative pay after the change were more likely to quit, whereas workers with a higher relative pay did not change their behavior. These important studies are informative

[5]However, individuals might accept or demand inequality on the grounds of equity (Konow, 2000). For example, wage differentials could be viewed as equitable if they reflect observable productivity differentials (Abeler et al., 2010; Breza et al., 2018). This could even be the case if differences in productivity are possible, but unobserved (Charness and Kuhn, 2007).

about the effect of wage inequality on labor supply decisions. However, they do not examine the effect of wage differences that are brought about by unfair chances or discrimination.

II.B. Unfair Chances

Economists have long considered the welfare implications of assessing inequality in terms of ex ante chances and ex post outcomes (Harsanyi, 1955; Diamond, 1967; Hammond, 1981; Epstein and Segal, 1992; Fleurbaey, 2010). Models of social preferences have also recently incorporated a dislike for unequal chances, usually referred to as a concern for ex ante or procedural fairness (Karni and Safra, 2002; Bolton et al., 2005; Trautmann, 2009; Krawczyk, 2011; Saito, 2013). A number of laboratory experiments have lent empirical support to the notion that individuals take into consideration the fairness of chances when making distributive decisions (Bolton et al., 2005; Karni et al., 2008; Krawczyk and Le Lec, 2010; Krawczyk, 2010; Brock et al., 2013; Cappelen et al., 2013; Andreoni et al., 2016; Grimalda et al., 2016; Trautmann and van de Kuilen, 2016; Cettolin and Riedl, 2016; Miao and Zhong, 2018). However, how ex ante chances influence the labor supply decisions of workers has not been investigated.[6]

II.C. Gender Discrimination

Women face a gender gap in earnings, have lower promotion chances, are less present in high-paid jobs, work less hours, work more part time, and have a lower labor participation (for overviews, see, Altonji and Blank, 1999; Riach and Rich, 2002; Goldin, 2014; Blau and Kahn, 2017). A number of studies suggest that demand-side gender discrimination plays an important role in explaining women's disadvantaged labor market position (see, e.g., Bertrand and Duflo, 2017; Neumark et al., 1996; Goldin and Rouse, 2000; Sarsons et al., 2019).[7]

Interestingly, potential supply-side effects of discrimination are much less studied. Parsons et al. (2011) present evidence from American baseball showing that minority players change their behavior in response to discrimination by officials. Glover et al. (2017) show that ethnically-biased managers in a large French grocery chain decrease minority workers' productivity and labor supply. A field experiment of Ibañez and Riener (2018) examines some aspects of gender discrimination (Affirmative Action for women) on job applications. In these studies, discrimination changes the monetary incentives for workers who are discriminated, so that the response to discrimination is entangled with the change in incentives.

Studies in medicine and psychology show that discrimination is correlated with serious negative consequences for physical and mental well-being. In a meta-analytic review, Pascoe

[6] Organizational psychologists have studied a related concept referred to as procedural justice. According to one prominent form of procedural justice, procedures are fair to the extent that decisions are "consistent" and without "bias" (see, e.g., Leventhal et al., 1980; Skarlicki and Folger, 1997).

[7] In this literature, taste-based discrimination (Becker, 1971) and statistical discrimination (Phelps, 1972; Arrow, 1973) are the two most discussed forms of discrimination. Other forms include language discrimination (Lang, 1986), implicit discrimination (Bertrand et al., 2005), attention discrimination (Bartoš et al., 2016), and stereotyping (Bordalo et al., 2016).

and Smart Richman (2009) link discrimination to a range of psychological issues, such as anger, stress, anxiety, distress, and low general well-being, all of which we can reasonably expect to considerably lower one's work satisfaction. However, the reviewed studies do not investigate the impact on workers' labor supply decisions.

III. Theoretical Framework

We adapt the framework of Card et al. (2012) and Breza et al. (2018) to model how workers may react to unequal wages, unfair chances, and gender discrimination. In the original model, wage inequality between workers engaged in the same work decreases work satisfaction or morale, which translates into lower marginal utility from work and thus into lower labor supply. We extend this model to also account for chances in the process leading to unequal wages. That is, next to wage inequality, unfair chances are assumed to decrease marginal utility from work. To explore the role of gender discrimination, we assume that individuals are more averse to unfair chances caused by gender discrimination than to unfair chances coming from an unspecified source.[8] We use this model to derive most of our hypotheses, which are formulated in Section V.

Consider two workers, i and j, engaged in the same work receiving piece-rate wages w_i and w_j, which are known to both workers. There is no interaction between the two workers and they do not observe each other's labor supply. A worker, say i, chooses labor supply l_i by taking into account his or her own wage, the wage of the other worker j, the chances that lead to their respective wages, and the cost of providing labor.

The modeling of marginal disutility created by ex post wage inequality is inspired by Fehr and Schmidt (1999) and is also used in Breza et al. (2018). It is denoted P_i and given by

$$P_i(w_i, w_j) = \alpha_i \max\{w_j - w_i, 0\} + \beta_i \max\{w_i - w_j, 0\}, \tag{1}$$

where the first term on the right-hand side measures the marginal disutility from disadvantageous wage inequality and the second term the marginal disutility from advantageous wage inequality, with $\alpha_i > \beta_i > 0$.[9] That is, wage inequality produces a marginal disutility, and this marginal disutility is greater for disadvantageous inequality than for advantageous inequality.

The marginal disutility created by unfair chances, denoted by A_i for ex ante inequality, is inspired by Saito (2013)[10] and takes the form

$$A_i(Ew_i, Ew_j) = \alpha_i' \max\{Ew_j - Ew_i, 0\} + \beta_i' \max\{Ew_i - Ew_j, 0\}, \tag{2}$$

[8] Unlike Breza et al. (2018) and in line with our experimental implementation, we rule out moral hazard and assume that work effort is fully contractible.

[9] This assumption is based on the empirical evidence reported in the studies investigating the effect of wage inequality on labor supply (Bracha et al., 2015; Breza et al., 2018; Dube et al., 2019), which suggests that, on average, $\alpha_i > \beta_i = 0$ or $\alpha_i > \beta_i > 0$. The literature on social preferences often makes the weaker assumption $\alpha_i \geq \beta_i \geq 0$ (e.g., Fehr and Schmidt, 1999). In our model, if $\alpha_i = \beta_i > 0$, then the effect of unequal wages on marginal disutility is the same for advantaged and disadvantaged workers. If $\alpha_i = \beta_i = 0$, then the problem collapses to standard selfish preferences and unequal wages do not affect the morale of workers.

[10] For earlier theoretical work combining social preferences and the effect of (un)fair chances, see Bolton et al. (2005) and Trautmann (2009).

where Ew_i and Ew_j denote expected wages. Similar to equation (1), here the first term on the right-hand side reflects the marginal disutility from disadvantageous expected wage inequality, and the second term that from advantageous expected wage inequality. As above we assume that $\alpha_i' > \beta_i' > 0$ but we allow for $\alpha_i \neq \alpha_i'$ and $\beta_i \neq \beta_i'$. That is, the disutility weights placed on wage inequality and unfair chances may differ.

We embed the aversion to unequal wages and the aversion to unfair chances described in equations (1) and (2) in the labor supply decision in the following way. A worker i chooses labor supply l_i in order to maximize the utility function

$$U_i(w_i, w_j, l_i) \;=\; w_i l_i \;-\; P_i(w_i, w_j) l_i \;-\; A_i(Ew_i, Ew_j) l_i \;-\; \frac{l_i^2}{2}. \tag{3}$$

In equation (3), the first term on the right-hand side corresponds to the utility of monetary earnings derived from working, the second term is the disutility created by wage inequality, and the third term reflects the disutility created by unfair chances. The final term is the utility cost of providing labor.[11]

Assuming an interior solution, the optimal labor supply is given by

$$l_i^* \;=\; w_i \;-\; P_i(w_i, w_j) \;-\; A_i(Ew_i, Ew_j). \tag{4}$$

Given our assumptions about $\alpha_i, \beta_i, \alpha_i'$ and β_i', the term $P_i(w_i, w_j)$ is strictly positive when wages are unequal and the term $A_i(Ew_i, Ew_j)$ is strictly positive when chances are unfair. Thus, unequal wages and unfair chances both reduce the optimal labor supply.[12] Disadvantageous inequality reduces the optimal labor supply more than does advantageous inequality because of $\alpha_i > \beta_i$ and $\alpha_i' > \beta_i'$.

Regarding gender discrimination, we posit that it translates into further marginal disutility from unequal wages or unfair chances. As before, we assume that the marginal disutility caused by discrimination is greater for disadvantaged than for advantaged workers. That is, gender discrimination in chances would increase α_i' and β_i', and would increase α_i' more than β_i'. Therefore, unfair chances based on gender discrimination reduce the optimal labor supply more than unfair chances based on an unspecified source and the labor supply reduction is greater for disadvantaged than for advantaged workers.

IV. Design of Study 1

IV.A. Experiment

We hired workers on an online labor platform to perform a real effort task, which consisted of entering lines of random characters. Each worker was assigned the same task and carried it out individually, entering one line at a time. The payment was on a piece-rate basis, that is, a worker

[11] For simplicity, we assume a quadratic cost function. The hypotheses derived from the model stay qualitatively the same when assuming any other strictly increasing and strictly convex cost function.

[12] Equation (4) assumes an interior solution. If the optimal labor supply is a corner solution—zero or maximum labor supply—altering the inequality of wages or the unfairness of chances may not affect the optimal labor supply.

received a fixed payment per correctly entered line. If a mistake was made when entering a line, the worker was informed and had to correct it before proceeding to the next line. The length of the lines increased with the number of lines completed, which made the task increasingly harder over time.[13] Each worker decided individually how many lines to enter. A worker could stop working at any time by leaving the experiment. Workers were informed about this and that they could not reenter the experiment once they had left. They were also instructed that they could work for at most 65 minutes. The number of lines entered is our measure of labor supply.

Table 1: Study 1—Wages and Chances of Two Workers in a Pair for each Payment Scheme

Payment Scheme	Wage of Worker, Wage of Other Worker	Chance of Worker, Chance of Other Worker*	Source of Chances
EQLOW	£0.03, £0.03	-	-
EQHIGH	£0.06, £0.06	-	-
UNEQFAIR	£0.03, £0.06	50%, 50%	Unspecified
UNEQUNFAIR	£0.03, £0.06	25%, 75%	Unspecified
UNEQDISCR	£0.03, £0.06	25%, 75%	Gender Discrimination

Note: * Chances describe the probability of receiving the high wage (£0.06).

We implemented five different payment schemes (our treatments), which are summarized in Table 1. Each worker was randomly assigned to one payment scheme and anonymously paired with another worker in the same scheme engaged in the same task on the platform. The schemes determine, within a worker pair, the possible wages and chances leading to these wages. In the schemes EQLOW and EQHIGH, no chances are involved and both workers receive either the low piece-rate wage of £0.03 or the high piece-rate wage of £0.06. These treatments serve as controls for the labor supply effect of receiving a low or a high wage when inequality in wages and chances is absent.

In the three other schemes, the two workers in a pair face equal or unequal chances to obtain the high or the low wage. Wages are randomly drawn such that one worker receives the high wage and the other worker receives the low wage. In UNEQFAIR both workers have a fair chance of 50% to receive the high wage, whereas in UNEQUNFAIR one worker faces a low chance of 25% and the other worker faces a high chance of 75% to receive the high wage. In both of these schemes workers do not receive a reason for why the chances are allocated in this manner. In contrast, in the scheme UNEQDISCR the unfair chances explicitly discriminate one gender over the other. That is, one worker in the pair is informed that she (he) faces a 25% chance of receiving the high wage because she (he) is a woman (man), and that the other worker faces a 75% chance of receiving the high wage because he (she) is a man (woman). In this treatment, in half of the cases men face higher chances and in the other half women face higher chances.[14]

[13]The number of characters contained in a line ranged from 10 at the start to 26 at the end. We implemented this to mimic an increasing and convex cost of labor supply. There was a maximum of 85 lines and workers were not informed about this beforehand (see Figure A1 of the appendix for two screenshots of the task). Further details on the task can be found in the instructions of the experiment, which are provided in Appendix C.

[14]Workers were not informed about this balance.

Each worker in a pair was first informed about their payment scheme, followed by information about their own resulting wage and the wage of the other worker. To emphasize the piece-rate nature of the wage, it was described as a "payment per line."[15]

At the start of the experiment, a worker electronically signed an informed consent form and then read the instructions. Each worker had to correctly answer nine comprehension questions and go through a practice phase to become familiar with the task. Only thereafter was the worker assigned to a payment scheme. After having learned his or her own wage and the wage of the other worker in the pair, a worker could start working on the task. Thus, workers only started working after any uncertainty about their own and the other worker's wage was resolved.

IV.B. Online Labor Markets

We recruit workers on the UK-based online platform Prolific (www.prolific.ac). The use of such online labor markets for experiments has gained in popularity among economists in recent years.[16] For our research the use of an online platform provides important advantages over a laboratory experiment or a field experiment inside a firm, but it also has some potential shortcomings. In what follows we discuss these advantages and how we deal with potential disadvantages.

The first advantage is that the online platform greatly reduces the possibility of peer effects, because workers on the platform do not interact with each other in any way during the task and can quit working without other workers noticing it. On the platform there is no channel through which workers could communicate with each other and to our knowledge also no informal website exists through which this happened. There are also restrictions on accounts and on participation per IP address, which we discuss below. Moreover, only a subset of registered potential workers meeting our criteria are invited to participate, which limits the probability that registered participants who know each other are invited.[17] In a post-experiment questionnaire,

[15]In the appendix, Figure A3(a) provides a screenshot example of what workers saw when they were informed about the payment scheme, and Figure A3(b) shows a screenshot example of what they saw when they were informed about the resulting wages.

[16]Examples include Pallais (2014) on inexperienced workers, Kuziemko et al. (2015) on redistribution preferences, Gilchrist et al. (2016) on employer-employees relationships, Pallais and Sands (2016) and Horton (2017a) on labor market referrals and recommendations, Bordalo et al. (2016) on stereotypes, Horton (2017b) on minimum wages and employment, Lyons (2017) on diversity and production in teams, Coffman et al. (2016) on anti-gay sentiments, Coffman et al. (2017) and Sarsons et al. (2019) on gender discrimination, and De Quidt et al. (2018) on experimenter demand effects. Horton et al. (2011), Arechar et al. (2018), and Snowberg and Yariv (2018) report that common economic games and elicited behavior in online and laboratory experiments provide qualitatively similar results. Bohren et al. (2018) also use an online scientific platform to study gender discrimination, where individuals are not paid but volunteer. Furthermore, several studies use existing labor data from online platforms, e.g., Ghani et al. (2014), Stanton and Thomas (2015), and Dube et al. (2018). Finally, see Chen and Konstan (2015) for a survey of several experiments on different platforms.

[17]Our criteria were: UK is the country of residence, registration as a man or woman, and an approval rate of at least 80% for previous participation in other studies. The number of registered individuals meeting our criteria was greater than 6,000. Workers meeting the selection criteria could register for our experiment without receiving an invitation email if they logged in on the website and selected our experiment, provided that our required number of workers had not been attained. The fact that our experiment was almost fully conducted within 24 hours limits this possibility. Importantly, there is no gain for workers from discussing or working with someone else during our experiment due to the nature of the task.

95.4% of workers reported that they did not discuss the task with someone else when deciding whether or not to participate, and a similar number (97.4%) declared that they completed the task without the help of someone else. In a field experiment inside a firm, in contrast, workers might communicate with others, observe how much others work, and news might spread that there exist different payment schemes. Also, in the laboratory it is likely that participants receive cues about the behavior of others because the typical experiment has multiple participants inside the same room.[18]

The second advantage is that it enables us to recruit a relatively large number of workers at reasonable costs, which increases the statistical power to detect differences in labor supply across payment schemes and increases the robustness of results (Camerer et al., 2016). The online platform also provides access to a pool of workers with more diverse demographic backgrounds than a sample of undergraduate students. The third advantage is that, with approval from an ethical review committee, we were allowed to engage in gender discrimination on the platform. This would have been difficult to implement in a field experiment inside a firm for legal and other reasons.

We took precautions to minimize potential problems that are associated with conducting experiments using online platforms. As participants have to read and understand the instructions of the experiment without support, it might be that they do not read them carefully enough or do not fully understand them. To ensure proper reading and understanding, we required participants to correctly answer nine exhaustive comprehension questions about the instructions. Participants who failed a question three times were automatically excluded and did not participate in the experiment. Another issue might be that participants do not fully trust that the instructions are truthful because online platforms do not necessarily have the reputation to be deception-free. To minimize this possibility, we made clear in the invitation to the experiment and again in the instructions that we do not use deception and that this is the standard in economic experiments. In a post-experiment questionnaire, participants report that they understood the instructions well and that they largely trusted that the instructions were truthful.[19]

Another concern might be that workers participate more than once, because it is not possible to directly verify the identity of participants. To minimize this possibility, the platform employs a number of measures to prevent duplicate accounts. These measures include limit-

[18] We have considered to run the experiment in the laboratory, but decided against it because it would have been very difficult and expensive to avoid peer effects. For instance, any worker who stops working and leaves the laboratory is likely to be noticed by other workers. An alternative would have been to have only one worker at a time in the laboratory, but this causes at least three problems: (1) it is extremely time consuming to collect a large enough number of observations, (2) it may open the door to session effects (Fréchette, 2012), and (3) information regarding the experiment can spread among potential participants, because the experiment would take place over a long time period.

[19] The comprehension questions also prevented automatic programs (robots) from entering our experiment by passing as human workers. Importantly, exclusion of subjects who failed was independent of the payment schemes, because they were excluded from the experiment before they were allocated to a scheme. Participants reported how well they understood the instructions on a Likert scale from 1 (Not at All) to 7 (Very Well) with a mean answer of 6.27 (SD = 1.04, N = 1,254) and reported how much they trusted the instructions on a Likert scale from 1 (Not at All) to 7 (Completely) with a mean answer of 5.94 (SD = 1.42, N = 1,254).

ing participation in an experiment to once per account, limiting the number of accounts per IP address, limiting participation in an experiment to once per IP address, requiring a unique non-voice over IP phone number per account, and limiting accounts to one per Paypal or Circle account for payment. The platform also forbids the use of VPNs and tracks changes in the country of connection and other suspicious participation patterns.[20]

IV.C. Procedures and Demographics

Workers on the online platform were invited via email in January 2018 and freely decided to participate by logging in the experiment website.[21] Not everyone who logged in actually participated. Specifically, the sample does not include the following individuals. First, the software automatically prevented participation of 281 individuals who did not complete the comprehension questions. Second, eight individuals chose to quit themselves at the end of the comprehension questions or during the practice phase. Third, 48 individuals were excluded because they exceeded the time limit.[22] Fourth, seven individuals where removed because the reported gender in the experiment did not correspond to their gender in the platform database.

In total, our sample consists of 1,271 workers who successfully completed the comprehension questions and participated in the experiment. On average, those workers spent 26.35 (SD = 15.56) minutes in the experiment, and were paid 2.59 (SD = 1.53) pounds.[23]

V. Hypotheses[24]

All hypotheses refer to workers who do not beat the odds. That is, *low*-wage workers in the payment scheme without chances or with fair or low chances to receive the high wage in payment schemes with chances, and *high*-wage workers in the payment scheme without chances or with fair or high chances to receive the high wage in payment schemes with chances. We only consider these workers because too few workers beat the odds for an informative statistical analysis.

[20]Our study was registered using two separate experiments on the platform, one only accessible to men and one only accessible to women. This is a feature of the platform, which requests that filtering by gender be done in this manner. This means that participation in the study was limited to one man and one woman per IP address (using gender reported on the platform). A total of 6% of participants had the same IP address at the time of their participation as another participant of the other gender. The address can be the same for different reasons, e.g., workers participate from the same house, public space or workplace. Note that Paypal and Circle also take steps to prevent duplication of accounts.

[21]The invitation email can be found in Section C of the appendix.

[22]These individuals went over the time limit of 65 minutes despite being explicitly forbidden from doing so in the experiment description on the platform, and being provided with a time countdown from 65 to 0 minutes during their work to remind them of the time limit.

[23]Table A1 of the appendix summarizes participants' demographic characteristics. It also reports F-tests showing that there are no significant differences in characteristics across payment schemes.

[24]Our hypotheses were preregistered before the execution of the experiment: American Economic Associations Randomized Control Trials Registry (ID: AEARCTR-0002655).

Table 2: Study 1—Predicted Labor Supply in Each Payment Scheme

Payment Scheme	Low-Wage Worker i	High-Wage Worker i
EQLOW/EQHIGH	$l_i^E(w_l) = w_l$	$l_i^E(w_h) = w_h$
UNEQFAIR	$l_i^F(w_l) = w_l - P_i(w_l, w_h)$	$l_i^F(w_h) = w_h - P_i(w_h, w_l)$
UNEQUNFAIR	$l_i^U(w_l) = w_l - P_i(w_l, w_h) - A_i(Ew_l, Ew_h)$	$l_i^U(w_h) = w_h - P_i(w_h, w_l) - A_i(Ew_h, Ew_l)$
UNEQDISCR	$l_i^D(w_l) = w_l - P_i(w_l, w_h) - A_i^D(Ew_l, Ew_h)$	$l_i^D(w_h) = w_h - P_i(w_h, w_l) - A_i^D(Ew_h, Ew_l)$

Note: The predicted labor supply of a worker i is given by Equation (4) in Section III: $l_i^* = w_i - P_i(w_i, w_j) - A_i(Ew_i, Ew_j)$, where the term $A_i^D(Ew_i, Ew_j)$ indicates the presence of gender-discriminatory chances ($A_i^D(Ew_i, Ew_j) > A_i(Ew_i, Ew_j)$). Predictions are for workers who do not beat the odds.

In what follows, the first three hypotheses are based on the optimal labor supply derived in the theoretical framework presented in Section III.[25] Table 2 shows the predicted labor supply in each payment scheme for low- and high-wage workers. Recall that, for unequal wages and unfair chances, P_i and A_i take on positive values, and for unfair chances, A_i takes on smaller values than A_i^D. From the table, it then follows straightforwardly that, for a low-wage worker i,

$$l_i^E(w_l) > l_i^F(w_l) > l_i^U(w_l) > l_i^D(w_l),$$

and for a high-wage worker i,

$$l_i^E(w_h) > l_i^F(w_h) > l_i^U(w_h) > l_i^D(w_h).$$

This leads to our first two hypotheses.

HYPOTHESIS 1, LOW-WAGE WORKERS: For low-wage workers, labor supply ranks across payment schemes as follows: EQLOW > UNEQFAIR > UNEQUNFAIR > UNEQDISCR.

HYPOTHESIS 2, HIGH-WAGE WORKERS: For high-wage workers, labor supply ranks across payment schemes as follows: EQHIGH > UNEQFAIR > UNEQUNFAIR > UNEQDISCR.

In the model, we assume that being in a disadvantageous position (low wage or low chance of receiving the high wage) creates a larger marginal disutiliy from inequality than being in an advantageous position (high wage or high chance of receiving the high wage).[26] We also assume that gender-discriminatory unfair chances are disliked more than unfair chances from an unspecified source, and that negative discrimination is worse than positive discrimination. This implies that, at a given wage, unequal wages, unfair chances, and gender discrimination each decrease the optimal labor supply more for a low-wage worker than for a high wage

[25] The theoretical optimal labor supply is based on the interior solution. We conducted a pilot study in advance to ensure that the parameters of the experiment (e.g., piece-rate wages, length of lines, duration) do not produce too many corner outcomes where workers do not work at all or finish all tasks. As we will see in Section VI, workers in the experiment indeed overwhelmingly choose an interior outcome.

[26] See Section III, Equations (2) and (3) together with $\alpha_i > \beta_i$ and $\alpha_i' > \beta_i'$.

worker. More formally, it holds that[27]

$$
\begin{aligned}
l_i^E(w_l) - l_i^F(w_l) &> l_i^E(w_h) - l_i^F(w_h), \\
l_i^F(w_l) - l_i^U(w_l) &> l_i^F(w_h) - l_i^U(w_h), \text{ and} \\
l_i^U(w_l) - l_i^D(w_l) &> l_i^U(w_h) - l_i^D(w_h).
\end{aligned}
$$

Our third hypothesis is therefore as follows.

HYPOTHESIS 3, DISADVANTAGE VS. ADVANTAGE: For each of the following comparisons, the labor supply decrease is greater for low-wage workers than for high-wage workers: UNEQFAIR vs. EQLOW/EQHIGH, UNEQUNFAIR vs. UNEQFAIR, and UNEQDISCR vs. UNEQUNFAIR.

As gender discrimination is generally experienced by women rather than by men in society, discriminating against women might have a different effect on labor supply than discriminating against men. However, a priori the direction of the difference is unclear. On the one hand, discrimination against women worsens existing inequalities and may be especially painful for them, and men might perceive discrimination against them as a justified compensation for everyday discrimination of women. This might lead to a strong negative labor supply reaction by women but a positive or neutral one by men. On the other hand, women may have weaker negative labor supply reactions because they are used to discrimination, and men may be habituated to higher chances so that their new experience of lower chances may be especially frustrating and thus strongly decrease their labor supply. Therefore, our fourth hypothesis, which concerns this possible gender difference, is not directed and we state the null hypothesis.

HYPOTHESIS 4, GENDER AND NEGATIVE DISCRIMINATION: The difference in labor supply between UNEQUNFAIR and UNEQDISCR is equal for both genders.

VI. Results

In this section we first report descriptive statistics regarding labor supply under the different payment schemes followed by tests of our four hypotheses. Thereafter, we present some exploratory analyses on potential gender differences beyond our hypotheses. As mentioned above, we measure labor supply as the number of lines completed, which ranges from 0 to 85.

VI.A. Descriptive Statistics

Figure 1 summarizes labor supply by type of worker and payment scheme.[28] For low-wage workers, the scheme EQLOW generates the largest mean labor supply, followed by UNEQUN-

[27]To obtain these three inequalities we use Table 2. For the first inequality we have $l_i^E(w_l) - l_i^F(w_l) = P_i(w_l, w_h)$ and $l_i^E(w_h) - l_i^F(w_h) = P_i(w_h, w_l)$. Because $\alpha_i > \beta_i$, it holds that $P_i(w_l, w_h) > P_i(w_h, w_l)$. The two other inequalities are obtained in the same manner.

[28]More detailed descriptive statistics are reported in the appendix, Table A2.

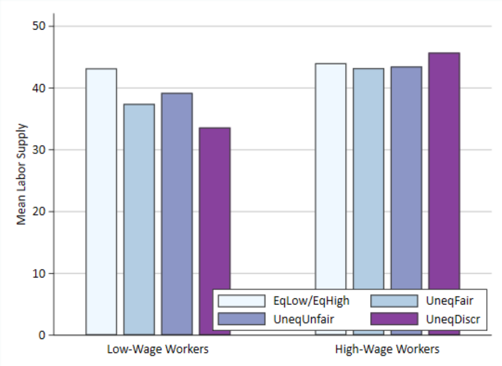


Figure 1: Study 1—Mean Labor Supply per Payment Scheme
Note: Mean labor supply, low-wage workers: 43.20 (EQLOW), 37.44 (UNEQFAIR), 39.22 (UNEQUNFAIR), 33.62 (UNEQDISCR). Mean labor supply, high-wage workers: 44.03 (EQHIGH), 43.23 (UNEQFAIR), 43.50 (UNEQUNFAIR), 45.74 (UNEQDISCR). *N* ranges from 127 to 145 workers per payment scheme.

FAIR and UNEQFAIR (with little difference between them), and then by UNEQDISCR. For high-wage workers, all schemes produce a comparable mean labor supply.[29]

VI.B. Test of Hypotheses

To test our hypotheses we employ non-parametric rank tests as well as Tobit regressions to account for lower-bound and upper-bound censoring of the dependent variable. For the few cases where the two techniques lead to different results in terms of statistical significance, we give priority to the non-parametric tests.[30] Table 3 shows coefficient estimates of the Tobit regressions of labor supply on dummies for the payment schemes, separately for low-wage workers and high-wage workers. EQLOW and EQHIGH serve as the respective reference schemes.

To test Hypotheses 1 and 2, we conduct pairwise comparisons between the schemes using non-parametric Dunn's tests and regression estimates.[31] We use one-sided tests because our

[29]A comparison between low and high wages shows a significantly larger labor supply when a high wage is paid only if gender discrimination is involved (EQLOW/EQHIGH, $p = 0.72$; UNEQFAIR, $p = 0.12$; UNEQUNFAIR, $p = 0.20$; UNEQDISCR, $p < 0.001$; two-sided t-tests). Mann-Whitney tests qualitatively return the same results. The implied wage elasticities of labor supply for our task are 0.02 in EQLOW/EQHIGH, 0.16 in UNEQFAIR, 0.11 in UNEQUNFAIR, and 0.36 in UNEQDISCR. Overall, these elasticities are in keeping with those estimated on online labor platforms. For instance, Dube et al. (2018) estimate the market-wide elasticity on Amazon Turk to be around 0.10. We also note that workers in our study are unlikely to expect immediate better-paying alternative work on the platform. This is because the platform prevents participants from observing any other work available before they finish a study and there is usually relatively little work available (few platform participants work more than once per day).

[30]We do so because non-parametric tests do not assume that error terms are normally distributed. For the Tobit regressions, we use robust standard errors because we find evidence of heteroscedasticity in our data. In the regressions, we also include control variables (age, gender, ethnicity, student status, employment status, experience on the platform, an index reflecting the percentage of approved participation in tasks on the platform, and day and time of participation).

[31]Dunn's test (Dunn, 1964) allows us to conduct multiple pairwise comparisons, and is considered to be the correct test after a Kruskal-Wallis (KW) test. Both non-parametric KW tests and parametric Wald tests (on the restriction that the three scheme coefficients from the regression are jointly equal to zero) confirm that labor supply differs across the four schemes for low-wage workers, but not for high-wage workers (low-wage workers: $p = 0.033$ KW, $p = 0.006$ Wald; high-wage workers: $p = 0.825$ KW, $p = 0.901$ Wald; two-sided tests).

Table 3: Study 1—Tobit Regressions of Labor Supply on Payment Schemes

	Low-Wage Workers	High-Wage Workers
Scheme	(1)	(2)
UNEQFAIR	−4.144	−1.309
	(4.337)	(4.547)
UNEQUNFAIR	−4.008	−1.607
	(4.115)	(4.397)
UNEQDISCR	−13.610****	1.146
	(4.051)	(4.220)
Controls	Yes	Yes
Prob > F	0.001	0.010
Pseudo R^2	0.011	0.008
N	533	542

Note: EQLOW and EQHIGH, respectively, serves as reference scheme. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$. Removing controls does not change the results qualitatively.

hypotheses are directional. Table 4 presents the p-values of the tests for the three main comparisons contained in each of the two hypotheses, with and without the Benjamini-Hochberg (BH) correction for multiple hypothesis testing within each hypothesis.[32]

Table 4: Study 1—P-values of Predicted Differences in Labor Supply between Payment Schemes

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EQLOW/EQHIGH >UNEQFAIR	0.017	0.170	0.050	0.255	0.371	0.387	1.000	1.000
UNEQFAIR >UNEQUNFAIR	0.878	0.513	0.878	0.513	0.780	0.473	1.000	0.709
UNEQUNFAIR >UNEQDISCR	0.021	0.006	0.031	0.019	0.886	0.752	0.886	0.752
N	539	533	539	533	543	542	543	542

Note: One-sided p-values in the direction predicted. BH corrections account for multiple hypothesis testing. Table A3 of the appendix presents results for all six possible pairwise comparisons; the results are the same.

We first discuss the labor supply responses of low-wage workers. Table 4 shows that for these workers, UNEQFAIR significantly decreases labor supply compared to EQLOW if we use non-parametric Dunn's tests, which is in line with Hypothesis 1. However, in contrast to this hypothesis, unfair chances have no additional negative effect, as labor supply in UNEQFAIR and UNEQUNFAIR are not significantly different. Finally, UNEQDISCR significantly reduces labor supply compared to UNEQUNFAIR, as predicted.[33]

[32]The BH correction (Benjamini and Hochberg, 1995) is a common False Discovery Rate procedure, which controls for the probability of false positives among significant results. It differs from Family-wise Error Rate procedures such as Holm-Bonferroni (Holm, 1979), which control for the probability of at least one false positive among significant results.

[33]To evaluate whether these results hold for the extensive or the intensive margin, we use a two-equation hurdle model with a lower bound. We find that relative to UNEQUNFAIR, the scheme UNEQDISCR reduces both the probability that low-wage workers start to work at all ($p = 0.026$ without and $p = 0.079$ with BH correction) and the labor supply of those who do decide to work ($p = 0.029$ without and $p = 0.086$ with BH correction). The negative effect on labor supply of UNEQFAIR relative to EQLOW appears to affect only the intensive margin

The detected effects are also economically significant, which can be assessed by comparing the mean labor supply across payment schemes. Unfair chances due to gender discrimination decreases mean labor supply from 43.20 in EQLOW to 33.62 in UNEQDISCR, corresponding to a reduction of 22%. Mean labor supply in UNEQUNFAIR is 39.22, indicating that gender-discriminatory unfair chances reduce mean labor supply by 15% compared to unfair chances from an unspecified source. Moreover, mean labor supply in UNEQFAIR is 37.44, representing a decrease of 13% compared to EQLOW.[34]

For high-wage workers, Table 4 reveals that none of the predicted inequalities in Hypothesis 2 hold. That is, high-wage workers provide similar labor supply across payment schemes. We summarize our first two results as follows.

RESULT 1, LOW-WAGE WORKERS: *(a) Gender-discriminatory unfair chances (*UNEQDISCR*) lower labor supply, compared to each other scheme. (b) Unfair chances from an unspecified source (*UNEQUNFAIR*) do not decrease labor supply compared to fair chances from an unspecified source (*UNEQFAIR*). (c)* UNEQFAIR *and* UNEQUNFAIR *reduce labor supply compared to* EQLOW.

RESULT 2, HIGH-WAGE WORKERS: *All payment schemes (*UNEQDISCR*,* UNEQUNFAIR*,* UNEQFAIR*,* EQHIGH*) produce similar labor supply.*

We now turn to Hypothesis 3, which states that the labor supply decrease is larger for low-wage workers than for high-wage workers, when comparing the payment schemes UNEQFAIR with EQLOW/EQHIGH, UNEQUNFAIR with UNEQFAIR, and UNEQDISCR with UNEQUNFAIR, respectively. We evaluate the hypothesis with a Tobit regression using dummy variables for UNEQFAIR, UNEQUNFAIR, and UNEQDISCR as well as their interactions with a dummy variable for high-wage workers. We also include a set of controls that is common to both low-wage and high-wage workers.[35]

Table 5 presents p-values from Wald tests conducted separately for each of the three inequalities that compose Hypothesis 3. It shows that the decreases in labor supply caused by UNEQFAIR compared to EQLOW/EQHIGH and by UNEQUNFAIR compared to UNEQFAIR are not significantly larger for low-wage workers than for high-wage workers, rejecting the first two inequalities of the hypothesis. However, as predicted, the labor supply reduction caused by UNEQDISCR relative to UNEQUNFAIR is (marginally) significantly larger for low-wage workers. Overall, our analysis provides evidence in favor of only the discrimination part of Hypothesis 3. We state our third result as follows.

(intensive margin: $p = 0.044$ without and $p = 0.087$ with BH correction; extensive margin: $p \geq 0.228$). In the appendix, Tables A10 and A11, respectively, show the results of the estimation and the p-values for the pairwise comparisons between payment schemes.

[34]In terms of pooled standard deviations, UNEQDISCR decreases labor supply by 0.35 standard deviations compared to EQLOW, and by 0.21 standard deviations compared to UNEQUNFAIR. The scheme UNEQFAIR reduces labor supply by 0.20 standard deviations relative to EQLOW. In their studies of labor supply responses to unjustified wage inequality in the relatively short term, Bracha et al. (2015) and Breza et al. (2018) find effect sizes for low-wage workers that are around 0.10 and 0.50 standard deviations.

[35]The regression results are reported in Table A4 of the appendix.

RESULT 3, DISADVANTAGE VS. ADVANTAGE: *(a) The decrease in labor supply caused by gender-discriminatory unfair chances (*UNEQDISCR*) relative to unfair chances from an unspecified source (*UNEQUNFAIR*) is larger for low-wage workers than for high-wage workers. (b) The decreases caused by unfair chances from an unspecified source (*UNEQUNFAIR*) relative to fair chances from an unspecified source (*UNEQFAIR*) and by* UNEQFAIR *relative to* EQLOW/EQHIGH *are similar for both types of workers.*

Table 5: Study 1—*P*-values of Predicted Differences in Labor Supply Effect of Payment Schemes between Low-Wage Workers and High-Wage Workers

	All Workers	
Predicted Inequality	(1)	(2)
Technique	Tobit	Tobit
BH Correction	No	Yes
UNEQFAIR × HighWage > 0	0.374	0.561
UNEQUNFAIR × HighWage > UNEQFAIR × HighWage	0.503	0.503
UNEQDISCR × HighWage > UNEQUNFAIR × HighWage	0.023	0.070
N	1075	1075

Note: One-sided *p*-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing.

Finally, we evaluate whether low-wage men and women respond differently to gender-discriminatory unfair chances (UNEQDISCR) relative to unfair chances coming from an unspecified source (UNEQUNFAIR), as stated in Hypothesis 4. We find that the mean labor supply of women differs considerably between the two schemes (42.37 in UNEQUNFAIR and 30.38 in UNEQDISCR; a difference of roughly 28%). In contrast, the mean labor supply of men is essentially equal in both schemes (36.13 in UNEQUNFAIR and 36.92 in UNEQDISCR).[36] Table 6 presents the estimates from a Tobit regression for low-wage workers in the two schemes with UNEQUNFAIR serving as the reference scheme, a dummy variable for women, and an interaction term of UNEQDISCR with the dummy variable for women. For men, discrimination does not significantly alter labor supply compared to unfair chances, as the coefficient of UNEQDISCR is insignificant ($p = 0.805$). However, the interaction term is negative and significant ($p = 0.040$), indicating that the labor supply decrease caused by gender-discriminatory chances relative to the same unfair chances without gender discrimination is stronger for women than for men.[37] Our fourth result is as follows.

RESULT 4, GENDER AND NEGATIVE DISCRIMINATION: *The decrease in labor supply caused by* UNEQDISCR *relative to* UNEQUNFAIR *is greater for low-wage women than for low-wage men.*

In summary, we find that in payment schemes involving chances to receive a low or a high wage, low-wage workers reduce labor supply relative to a scheme in which both workers

[36]Table A5 of the appendix presents more detailed descriptive statistics of labor supply by gender.
[37]The interaction term is also significant without controls ($p = 0.043$).

Table 6: Study 1—Tobit Regression of Labor Supply in UNEQUNFAIR and UNEQDISCR, for Low-Wage Men and Women

Scheme	Low-Wage Workers
UNEQDISCR	-1.363
	(5.510)
UNEQDISCR × Woman	-16.105^{**}
	(7.805)
Woman	1.736
	(6.077)
Controls	Yes
Prob $> F$	0.039
Pseudo R^2	0.012
N	283

Note: UNEQUNFAIR serves as baseline. Standard errors are indicated in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

receive the same low wage. This effect is especially pronounced when chances are gender discriminatory. Interestingly, there is no difference between fair and unfair chances when these come from an unspecified source. For high-wage workers, labor supply is unaffected by the nature of the payment scheme. We find evidence that the decrease in labor supply is stronger for low-wage workers than for high-wage workers, especially when gender-discriminatory chances are involved. Finally, women respond more emphatically to gender-discriminatory chances than do men.[38]

VI.C. Further Gender Differences

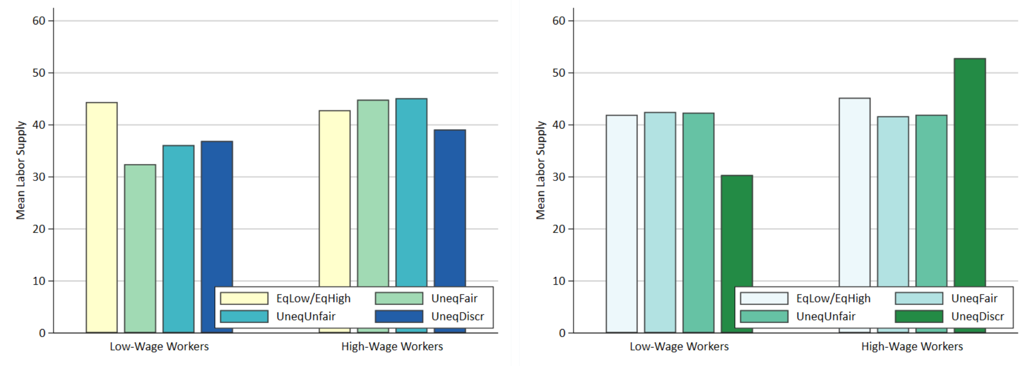
Here we report some additional analyses on gender differences that go beyond our pre-registered hypotheses. Figure 2 shows the mean labor supply of workers in each payment scheme, separately for men and women.[39] Focusing first on low-wage workers, we see that the labor supply reactions to the different types of payment schemes differ considerably between genders. Men lower their labor supply in response to any of the three payment schemes with unequal wages, whereas women decrease their labor supply only in response to unequal wages resulting from gender-discriminatory chances.

We test for differences in labor supply across payment schemes separately for men and women, using non-parametric Dunn's tests and Tobit regression estimates.[40] The tests corrob-

[38]In the appendix we provide supplementary analyses. Tables A12–A15 detail labor supply reactions to the payment schemes for four additional demographic groups: the young and the old, and the full-time and part-time employed or unemployed. Furthermore, our pre-registered measure of labor supply is the number of lines, but other measures of labor supply are conceivable. Tables A16 and A17 show workers' responses to the different payment schemes if we use time spent in the experiment as the labor supply measure. Note that this measure does not only include time spent working on the task. As such, it is most likely not an appropriate measure of labor supply. Nevertheless, the results are very similar to our pre-registered measure. Lastly, for completeness, Tables A18 and A19 describe how the relatively few workers who beat the odds respond to the payment schemes.

[39]More detailed descriptive statistics can be found in Table A5 of the appendix.

[40]For the regression, we use the same specification as in Table 3, but include a dummy variable for women



(a) Men

(b) Women

Figure 2: Study 1—Mean Labor Supply per Gender and Payment Scheme
Note: N ranges from 62 to 75 workers per payment scheme.

orate the impression conveyed by the descriptive statistics. Low-wage male workers significantly decrease their labor supply in UNEQFAIR compared to EQLOW ($p \leq 0.049$ after BH correction), but do not further decrease labor supply in response to the additional inequalities contained in UNEQFAIR and UNEQDISCR ($p \geq 0.358$ after BH correction). In contrast, low-wage female workers significantly decrease their labor supply only in UNEQDISCR relative to UNEQUNFAIR ($p \leq 0.009$ after BH correction) but not in response to other inequalities ($p \geq 0.457$ after BH correction). For high-wage workers the picture is quite different as they do not significantly reduce their labor supply in response to the different payment schemes, irrespective of their gender.

A closer look at the mean labor supply of high-wage women suggests an interesting pattern opposite to Hypothesis 2, namely an *increase* in labor supply in response to positive discrimination (see, Panel (b) of Figure 2). Dunn's tests indeed return significant differences when comparing women's labor supply in UNEQDISCR to their labor supply in the other three treatments (EQHIGH: $p = 0.055$, UNEQFAIR: $p = 0.007$, UNEQUNFAIR: $p = 0.012$; two-sided tests).[41]

VII. *Design of Study 2*

Study 1 strongly suggests that men and women differ in their response to the various forms of inequality embedded in the investigated payment schemes. It appears that low-wage men respond to wage inequality irrespective of the source of the inequality, whereas low-wage women reduce labor supply especially in response to gender discrimination. Study 1 also provides suggestive evidence that positive gender discrimination may increase the labor supply of high-wage women. Therefore, one aim of this second study is to better understand the different responses to discrimination by the two genders.

and interact this variable with the payment schemes. Tobit regression estimates can be found in Table A6 of the appendix. The p-values of all comparisons are reported in Table A7 and Table A8 of the appendix.

[41] Table A9 of the appendix reports p-values of comparisons and tests, using both Dunn's tests and Tobit regression.

Further, in Study 1 we implemented discrimination in an explicit way, in order to have full control over beliefs. However, arguably, gender discrimination in the field is more implicit and beliefs about whether or not discrimination is indeed the cause of the wage inequality between men and women may be an important factor driving labor supply responses. Hence, a second aim of this study is to examine whether explicit and implicit gender discrimination both affect labor supply and what role beliefs play.

To achieve these aims we implement new payment schemes with two features that distinguish them from those in Study 1. First, to explore the effect of different forms of discrimination, we implement an implicit discrimination payment scheme, next to an explicit discrimination payment scheme similar to the one in Study 1. In addition, to understand the role of beliefs we collect data on the extent to which participants believe that unequal payment in our experiment is due to gender discrimination. Second, since our emphasis is on the effect of different forms of discrimination on both genders, we implement unequal wages directly without chances.

VII.A. Design of the Experiment and Hypotheses[42]

The basic design is the same as for the experiment of Study 1. That is, workers receive instructions regarding the task and have to correctly answer a number of comprehension questions before they are informed about their payment scheme and can start working on the task. When workers receive information about the payment scheme, they are also informed that they are paired with another worker with whom they do not interact in any way. They do not receive any information about the other worker, except that in some treatments the gender of the other worker and the presence of discrimination is disclosed. All worker pairs are gender balanced.

We implement three new payment schemes, which are summarized in Table 7. In the baseline payment scheme, called UNEQ, one worker in a pair receives the high wage and the other worker receives the low wage. No explanation is given about why the wages are allocated in that way and workers do not receive information regarding the gender of the paired worker. Second, in the implicit discrimination payment scheme UNEQIM, next to the wage information, each worker is informed of the gender of the other worker in their pair. No further information is provided. Finally, in the payment scheme UNEQEX, each worker in the pair is explicitly informed that the high (low) wage is assigned based on gender. Assignment to payment schemes was gender balanced.

Our sample consists of 1593 workers who completed the experiment and none of them participated in Study 1.[43] On average, workers spent 34.10 (SD = 19.14) minutes, and were paid 2.96 (SD = 1.65) pounds.

[42]As for Study 1, the hypotheses were preregistered before the execution of the experiment: American Economic Associations Randomized Control Trials Registry (ID: AEARCTR-0003379).

[43]The sample was larger in each scheme than in Study 1, providing more statistical power to test for gender differences. The experiment was conducted in October 2018. We note a minor difference in the design relative to Study 1: here workers are allowed to work up to 70 instead of 65 minutes and the software automatically stops the experiment at 70 minutes. Table A20 of the appendix summarizes the participants' demographic characteristics. It also reports F-tests showing that there are no significant differences in characteristics across payment schemes.

Table 7: Study 2—Wages of the Two Workers in a Pair for each Payment Scheme

Payment Scheme	Wage of Worker, Wage of Other Worker	Gender Wage Gap
UNEQ	£0.03, £0.06	Unknown
UNEQIM	£0.03, £0.06	Implicit Gender Discrimination
UNEQEX	£0.03, £0.06	Explicit Gender Discrimination

Based on the theoretical framework and the results of Study 1 we derive three sets of hypotheses. The first set of hypotheses concerns low-wage workers who always face disadvantageous wage inequality.

HYPOTHESIS 5, LOW-WAGE WORKERS: *(5a) For men and women pooled, labor supply ranks across payment schemes as follows:* UNEQ > UNEQIM > UNEQEX. *(5b) For women, labor supply ranks in the same way as for the pooled sample. (5c) For men, labor supply does not respond to discrimination.*

Hypothesis 5a is based on the assumption of our theoretical model that discrimination can decrease labor supply through extra psychological costs. This implies the first inequality which finds empirical support in Study 1 (cf. Figure 1). The second inequality can be rationalized with two arguments. First, the psychological costs may be lower when discrimination is implicit than when it is explicit, and, second, a substantial share of workers in the implicit treatment may not attribute the wage inequality to discrimination and may thus not respond to it. Hypothesis 5b is implied by the same intuition and Hypothesis 5c is based on the observed responses of low-wage men in Study 1 (cf. Figure 2).

The second set of hypotheses concerns high-wage workers who always face advantageous wage inequality.

HYPOTHESIS 6, HIGH-WAGE WORKERS: *(6a) For men, labor supply ranks across payment schemes as follows:* UNEQ > UNEQIM > UNEQEX. *(6b) For women, labor supply ranks across payment schemes as follows:* UNEQ < UNEQIM < UNEQEX.

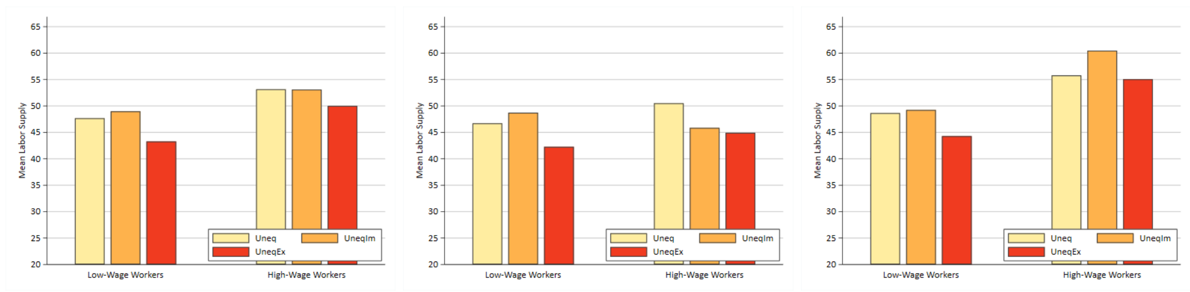
Hypothesis 6a states that men are averse to positive discrimination and thus in response decrease their labor supply, whereas Hypothesis 6b says that women favor positive discrimination and thus increase labor supply when facing it. Both hypotheses are based on the observed labor supply responses by men and women to explicit positive discrimination in Study 1 (cf. Figure 2), together with the idea that explicit discrimination has a stronger effect than implicit discrimination.

The final set of hypotheses makes statements regarding the difference in labor supply responses between men and women for both negative and positive discrimination.

HYPOTHESIS 7, DIFFERENCE WOMEN AND MEN: *Relative to no discrimination (*UNEQ*), (7a) in the joint schemes that involve negative discrimination (*UNEQIM *and* UNEQEX*), low-wage women decrease their labor supply more than do low-wage men, and (7b) in the joint schemes that involve positive discrimination, high-wage women increase their labor supply more than do high-wage men.*

Hypotheses 7a and 7b are both informed by the results of Study 1 and also implied by the combination of Hypotheses 5b and 5c, and Hypotheses 6a and 6b, respectively.

VII.B. Descriptive Statistics and Test of Hypotheses



(a) All Workers (b) Men (c) Women

Figure 3: Study 2—Mean Labor Supply per Payment Scheme

Note: Per payment scheme, N ranges from 265 to 266 workers in Panel (a), and from 132 to 134 workers in Panels (b)–(c).

Figure 3 gives an overview of the mean labor supply in each payment scheme, for low-wage and high-wage workers separately. Panel (a) provides the means pooled across genders, Panel (c) includes only women, and Panel (b) includes only men.[44]

We start with Hypothesis 5a, which concerns the behavior of low-wage workers pooled across genders.[45] Figure 3a suggests that there is little difference between schemes UNEQ and UNEQIM, whereas labor supply appears to decrease in scheme UNEQEX. Statistical tests corroborate this impression. Table 8 provides p-values of tests for the hypothesized changes in labor supply across schemes. It shows that labor supply does not differ significantly between UNEQ and UNEQIM. However, in UNEQEX labor supply is significantly reduced relative to UNEQIM. The effect of this latter comparison is large in economic terms: average labor supply is contracted by almost 12% in UNEQEX relative to UNEQIM. Labor supply is also significantly smaller in UNEQEX than in UNEQ ($p \leq 0.029$) and the effect amounts to a 9% reduction in labor supply.[46] This provides the following result, partially supporting our first hypothesis.

[44]Detailed descriptive statistics are reported in the appendix, Tables A21 and A22. Comparing high- and low-wage workers pooled across genders shows that labor supply is significantly larger for the former than for the latter in UNEQ ($p = 0.03$) and UNEQEX ($p = 0.01$) and close to marginally significant in UNEQIM ($p = 0.11$; all two-sided t-tests). Mann-Whitney tests qualitatively return the same results. The implied wage elasticities of labor supply are 0.11 in UNEQ, 0.08 in UNEQIM, and 0.16 in UNEQEX. These elasticities are comparable to those observed in Study 1, and are again in the typical range for those estimated on online labor platforms (e.g., Dube et al., 2018).

[45]To test our hypotheses we employ the same statistical approach as in Study 1. That is, we use rank tests (Dunn's tests) as well as Tobit regressions. We rely on the non-parametric tests whenever it is possible to use both techniques and when they return qualitatively different results. For conciseness we only report the main results here and relegate the detailed regression analyses to the appendix.

[46]When comparing labor supply of high-wage workers pooled across genders, we find that UNEQEX decreases labor supply relative to the other schemes, at marginal significance levels without the BH correction. However, as we did not have a hypothesis for this comparison we do not further elaborate on it. In terms of pooled standard

RESULT 5.1, LOW-WAGE WORKERS POOLED ACROSS GENDERS: *Explicit gender-discriminatory wage inequality (*UNEQEX*) significantly reduces labor supply, compared to both implicit gender-discriminatory and gender-neutral wage inequality (*UNEQIM *and* UNEQ*). In the latter two, labor supply does not differ.*

Table 8: Study 2—*P*-values of Predicted Differences in Labor Supply between Payment Schemes

	Low-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)
Technique	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes
UNEQ > UNEQIM	0.843	0.616	0.843	0.616
UNEQIM > UNEQEX	0.004	0.012	0.008	0.023
N	796	778	796	778

Note: One-sided p-values in the direction predicted. BH corrections account for multiple hypothesis testing. Both KW and Wald tests show that labor supply differs across the three schemes (p = 0.038 KW, p = 0.041 Wald, two-sided tests).

Figures 3(b)–(c) give an impression of how female and male low-wage workers respond to the different payment schemes. As for the pooled data, they show for both genders that labor supply in UNEQEX is reduced relative to the two other schemes, whereas there is little difference between UNEQ and UNEQIM.

Table 9 summarizes test results regarding labor supply responses to the schemes separately for the two genders (Hypotheses 5b and 5c).[47] The lower-left panel of the table shows that for low-wage women implicit gender-discriminatory wage inequality (UNEQIM) has no significant effect relative to gender-neutral wage inequality (UNEQ), whereas explicit gender-discriminatory wage inequality (UNEQEX) (marginally) significantly reduces women's labor supply relative to UNEQIM.[48]

The upper-left panel of the table shows that the test for differences in labor supply across payment schemes for low-wage men is qualitatively similar to the one found for women. Relative to gender-neutral wage inequality, low-wage men do not adjust labor supply when wage inequality is implicitly gender-discriminatory, but do significantly reduce labor supply in response to explicit gender discrimination, relative to when discrimination is implicit.[49]

deviations, UNEQEX decreases labor supply by 0.19 standard deviations relative to UNEQIM, and by 0.15 standard deviations relative to UNEQ.

[47]With these smaller sub-samples two-sided Kruskal-Wallis and Wald tests do not reject equality of labor supply across the three schemes, for high- and low-wage workers. We nevertheless proceed with pairwise comparisons because for women we are interested in directional changes and these tests cannot be one sided when comparing more than two treatments, and because most p-values are close to marginal significance for men (for details, see Table A25 of the appendix).

[48]The difference is nearly (marginally) significant when comparing labor supply in UNEQEX with labor supply under gender-neutral wage inequality in UNEQ. For an overview of all pairwise comparisons, see Table A26 of the appendix.

[49]Comparing labor supply under explicit gender-discriminatory wage inequality and gender-neutral wage inequality shows a (close to) marginally significant difference (see Table A26 of the appendix).

Table 9: Study 2—*P*-values of Predicted Differences in Labor Supply between Payment Schemes, for Men and Women

	Low-Wage Workers					High-Wage Workers			
Predicted (In)equality	(1)	(2)	(3)	(4)	Predicted Inequality	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit		Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes		No	No	Yes	Yes
Men									
UNEQ = UNEQIM	0.300	0.660	0.300	0.660	UNEQ > UNEQIM	0.050	0.185	0.076	0.185
UNEQIM = UNEQEX	0.024	0.064	0.048	0.129	UNEQIM > UNEQEX	0.169	0.179	0.169	0.358
N	398	391	398	391		399	388	399	388
Women									
UNEQ > UNEQIM	0.783	0.450	0.783	0.450	UNEQ < UNEQIM	0.035	0.182	0.069	0.364
UNEQIM > UNEQEX	0.036	0.092	0.071	0.185	UNEQIM < UNEQEX	0.973	0.733	0.973	0.733
N	398	387	398	387		398	391	398	391

Note: One-sided *p*-values are presented in the direction predicted for directed hypotheses; two-sided *p*-values are used for low-wage men. BH corrections account for multiple hypothesis testing. Tobit regressions results are presented in Table A26.

RESULT 5.2, LOW-WAGE WORKERS WOMEN AND MEN: *For low-wage women and men, explicit gender-discriminatory wage inequality (*UNEQEX*) reduces labor supply, compared to both implicit gender-discriminatory and gender-neutral wage inequality (*UNEQIM *and* UNEQ*). For both genders, implicit gender discrimination does not affect labor supply relative to gender-neutral wage inequality.*

Thus, our hypothesis regarding the effect of discrimination on women is partly supported (Hypothesis 5b). For men our hypothesis was that they will not respond at all to discriminatory wage inequality. We do, however, find a similar effect for both genders and, therefore, Hypothesis 5c is rejected.

We now turn to Hypotheses 6, regarding high-wage workers. We see from the upper-right panel of Table 9 that for high-wage men UNEQIM decreases labor supply relative to UNEQ, at (marginal) significance levels. The lower-right panel of the table shows for high-wage women that labor supply (marginally) significantly increases in UNEQIM compared to UNEQ. However, unlike in Study 1, explicit positive discrimination UNEQEX of women does not increase their labor supply, neither relative to UNEQIM nor relative to UNEQIM. Moreover, UNEQEX and UNEQIM produce similar labor supply for high-wage men.[50] We summarize in the following result which partly supports Hypothesis 6.

RESULT 6, HIGH-WAGE WORKERS WOMEN AND MEN: *For high-wage women, there is some evidence that implicit discrimination increases labor supply, whereas explicit discrimination does not. For high-wage men, both types of discrimination similarly reduce labor supply relative to non-discriminatory high wages.*

Moreover, regarding Hypothesis 7 we find that, relative to UNEQ, the joint two schemes involving *negative* discrimination do not decrease labor supply significantly more for women than for men (Wald test on restriction with Tobit regression coefficients, one-sided $p = 0.508$).

[50] See Table A26 for all pairwise comparisons.

In contrast, the joint two schemes involving *positive* gender discrimination do induce a larger decrease in labor supply for men than for women, at marginal significance levels (Wald test on restriction with Tobit regression coefficients, one-sided $p = 0.053$).[51] We therefore find support for Hypothesis 7b, but not for Hypothesis 7a.

VII.C. The Role of Beliefs in the Presence of Gender Discrimination

We have seen that negative gender discrimination adversely affects labor supply when discrimination is explicit but not when it is implicit (Results 5.1 and 5.2). An important distinctive feature of the explicit discrimination scheme is that there is no doubt about the discrimination and thus beliefs about discrimination are fixed. This is different for the implicit discrimination scheme where gender discrimination can only impact workers who perceive it as such. Therefore, a possible explanation for finding no overall effect of implicit gender discrimination is that sufficiently many workers did not believe (strongly enough) that their low wage was due to negative discrimination. Here we shine some light on the relationship between belief about the presence of negative gender discrimination and labor supply.

In the post-experiment questionnaire we asked all participants the extent to which they believed that gender discrimination was used in determining the low and the high wage.[52] Table 10 shows descriptive statistics on how much workers believed that discrimination was used for all three payment schemes, pooled as well as separately for men and women. As one would expect, when gender was not mentioned at all (UNEQ), the belief that gender discrimination was used was basically absent. In contrast, that belief increased when discrimination was implicit (UNEQIM), and it was strongest when discrimination was explicit (UNEQEX). The differences across treatments are highly significant (KW test for the restriction that all treatments generate the same belief and pairwise two-sided Dunn's tests: $p < 0.001$). A comparison between men and women shows that for each payment scheme women more strongly believed in the presence of discrimination than did men (UNEQ: $p = 0.041$, UNEQIM: $p = 0.009$, UNEQEX: $p < 0.001$; two-sided Mann-Whitney tests). Next we investigate the relationship between the belief in the use of negative gender discrimination and labor supply.

Table 10: Study 2—Belief that Gender Discrimination Determined Wages in the Experiment

	Low-Wage Workers								
	Both Genders			Men			Women		
	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*
UNEQ	1.56	1.13	259	1.45	1.07	130	1.67	1.18	129
UNEQIM	4.05	2.37	264	3.70	2.29	132	4.40	2.41	132
UNEQEX	5.38	2.09	265	4.99	2.18	133	5.78	1.92	132

Note: Belief scale that gender discrimination was used ranges from 1 (*Not at all)* to 7 (*Completely*).

Table 11 reports, for low-wage workers, results of tests on the relationship between the belief that gender discrimination determined wages in the experiment and labor supply, using

[51]The regression results are reported in Table A27 of the appendix.

[52]Workers answered the question "*During the task, did you believe that gender discrimination was used to determine your payment per line?*" on a Likert scale from 1 (*Not at all*) to 7 (*Completely*).

Spearman rank order tests and Tobit regressions, respectively. Specifications (1) and (2) show that for observations pooled across all three payment schemes and both genders, a highly significant negative correlation between belief in discrimination and labor supply exists. When focusing on the payment scheme with implicit gender discrimination, we find a (marginally) significant negative correlation between belief and labor supply. Qualitatively, the size of the effect is almost identical between the pooled schemes and UNEQIM alone. The table also shows that the negative effect of belief on labor supply appears to be mainly driven by women. While men's labor supply is unresponsive to the belief of being discriminated, the labor supply of women responds significantly to the belief of being negatively discriminated (specifications (5)–(6) and (7)–(8), respectively). These results strongly suggest that implicit negative gender discrimination having adverse effects on labor supply, for those who believe that they are discriminated.[53]

Table 11: Study 2—*P*-values of Negative Correlation between Low-Wage Workers' Labor Supply and Belief that Gender Discrimination Determined Wages in the Experiment

	ALL SCHEMES		UNEQIM					
	Both Genders		Both Genders		Men		Women	
Low-Wage Workers	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Spearman	Tobit	Spearman	Tobit	Spearman	Tobit	Spearman	Tobit
Correlation or Coefficient	−0.081	−1.774 to −1.246	−0.081	−2.492 to −1.502	0.004	0.287 to 1.079	−0.165	−4.750 to −3.154
p-value	0.012	0.005–0.018	0.096	0.032–0.100	0.963	≥0.495	0.029	0.003–0.028
N	788	771–788	264	260–264	132	130–132	132	130–132

Note: *p*-values are one-sided for both genders and for women (two-sided for men), reflecting the detrimental effect that we posit for negative discrimination in Study 2 for both genders pooled together and for women. Belief scale that gender discrimination was used ranges from 1 (*Not at all*) to 7 (*Completely*). We provide the range of *p*-values obtained from Tobit regressions with and without control variables, which are reported in Table A28 and A29.

VIII. Discussion

VIII.A. Unequal Wages and Labor Supply

Bracha et al. (2015), Breza et al. (2018), and Dube et al. (2019) (henceforth *BGL*, *BKS*, and *DGL*) study the effect of unequal wages on labor supply, but do not investigate the role of chances and discrimination. *BGL* conduct a laboratory experiment, *BKS* employ a field experiment, and *DGL* exploit a natural experiment. Here, we briefly compare our results concerning unequal wages resulting from fair chances or unfair chances with an unspecified source in Study 1 (UNEQFAIR and UNEQUNFAIR) to the results of these three studies.[54]

[53]Figures A2a–A2c in the appendix illustrate the labor supply effect of belief in negative discrimination using a median split. We also note that there is no significant correlation between the belief in facing positive discrimination and labor supply, when pooling payment schemes or in the implicit discrimination treatment scheme.

[54]*BGL* and *BKS* also consider cases where unequal wages may be justified, e.g., by productivity differences. We do not discuss these results.

In line with these studies, we find that unequal wages significantly decrease labor supply of low-wage workers. In addition, as do *BGL* and *DGL*, we find that unequal wages do not affect labor supply of high-wage workers. *BKS*, by contrast, report evidence that such workers may reduce their labor supply. They suggest that social tensions in workers' interactions drive the effect for high-wage workers, which could explain why those workers do not change their labor supply in our setting where workers cannot interact. Furthermore, regarding gender differences, we find that only low-wage men negatively respond to wage differences, whereas women do not. This is consistent with the result reported in *BGL* that only men respond to wage inequality.[55]

VIII.B. Fairness of Chances

One motivation of our research is to explore how the (un)fairness of initial chances between two workers affects their labor supply decisions once wages are known. To the best of our knowledge, no other study investigates this question. However, a number of scholars have analyzed the effect of ex ante fairness more generally. Closest to our research are the studies examining whether and how fairness of ex ante chances between individuals influences their equity judgments.

It has been shown that individuals are more likely to accept an unequal outcome that results from fair chances than one that results from unfair chances (e.g., Bolton et al., 2005; Grimalda et al., 2016). In stark contrast, we find that behavior of workers is insensitive to initial chances when they are generated by an unspecified source, as labor supply is almost identical under the payment schemes UNEQFAIR and UNEQUNFAIR.[56] However, workers in our study do respond to unfair chances coming from gender discrimination, a prominent form of procedural unfairness. This complements findings that workers' reaction to unequal wages depends on the reason behind wage inequality (e.g., workers may accept or even demand wage inequality if it reflects productivity differences, Abeler et al., 2010; Breza et al., 2018, or if some other justification is provided, Bracha et al., 2015).

VIII.C. Discrimination, Labor Supply, and the Gender Gap in Earnings

Goldin (2014) and Blau and Kahn (2017) report that the most important determinant of the modern gender earnings gap is that women exhibit lower labor supply. Women are less present in high-pay occupations, which usually demand long working hours (e.g., lawyer, manager, professor), and women work less and earn less *within the same occupation*, which typically offer rapidly-rising returns to working hours. The main explanation for the lower labor supply put forward in the literature is that women prefer temporal flexibility at work, notably working

[55] *BKS* only sample men and *DGL* do not report effects by gender. There are also other demographic differences between the four studies. For instance, *BGL* employ American university students, *DGL* use young American part-time workers, and *BKS* use Indian temporary workers who may be older. We use online UK workers with a median age of 36, most of whom are either full-time or part-time employed.

[56] We note that, as in the aforementioned experiments, in UNEQFAIR and UNEQUNFAIR chances are assigned without a specific reason.

less hours, because they have to bear a much greater share of household responsibilities (e.g., Bertrand et al., 2010; Flabbi and Moro, 2012; Goldin, 2014; Goldin and Katz, 2016; Wiswall and Zafar, 2017; Cortés and Pan, 2019).[57]

Our results that women decrease their labor supply in response to both gender-biased unfair chances and gender-discriminatory wages offer a complementary explanation for women's lower labor supply and lower earnings. Specifically, what our results suggest is that the discrimination experienced by women in labor markets may decrease their willingness to enter potential high-income occupations that demand long working hours and may also reduce their willingness to work long hours for a given salary within an occupation. In this case, the effect of discrimination on labor supply can become a self-fulfilling prophecy, and breaking this cycle may therefore prove to be no small feat.

Our results on the adverse effect of negative gender discrimination on labor supply also challenges an often-made assumption. Discrimination is commonly estimated by measuring the difference in earnings between men and women with the same characteristics—e.g., age, education, experience, hours worked and occupation—under the assumption that these characteristics are themselves unaffected by discrimination (see Oaxaca, 1973; Blinder, 1973, and Fortin et al., 2011 for an overview of the method). However, our results suggest that finding that men earn higher wages because they work longer hours can hide the fact that women work less hours (and incur disutility) exactly because of discrimination. That is, the impact of gender discrimination is likely underestimated.

We also provide limited evidence that positive gender discrimination may increase the labor supply of women. An exploratory analysis in our main study suggests that positive discrimination of women increases their labor supply. In our second study, which was partly designed to analyze this issue further, the implementation of implicit positive discrimination in favor of women also (marginally) increases their labor supply. However, explicit positive discrimination has no effect.[58] Interestingly, in Study 2 we also find that men react with a decrease in labor supply to positive gender discrimination in their favor. In our view these are interesting observations but it is too early to draw conclusions. We also note that, even if there is no direct effect on labor supply, a gender gap favoring women could send a valuable signal that there is no negative discrimination against women, which may increase their non-monetary marginal utility from work. We leave the examination of these issues for future research.

Finally, our studies also point to the important role that beliefs about gender discrimination may play for labor market outcomes. In both studies, in the payment schemes with explicit

[57] In a recent field experiment Mas and Pallais (2017) find that women indeed have greater preference for temporal flexibility, but that this is not enough to explain the gender earnings gap. There is evidence supporting other explanations for the modern gender earnings gap, including discrimination (e.g., Neumark et al., 1996; Goldin and Rouse, 2000; Reuben et al., 2014; Sarsons et al., 2019), differences in bargaining behavior and competitiveness (e.g., Niederle and Vesterlund, 2007; Buser et al., 2014; Card et al., 2015), differences in productivity (e.g., Mulligan and Rubinstein, 2008), social norms (e.g., Coffman, 2014; Bertrand et al., 2015; Bursztyn et al., 2017), and stereotypes (e.g., Bordalo et al., 2019).

[58] Previous studies have gathered support for beneficial effects of positive discrimination in other settings (Balafoutas and Sutter, 2012; Niederle et al., 2013; Ibañez and Riener, 2018), although they do not controls for monetary incentives as we do.

gender discrimination against women—where beliefs are fixed by design—there is an unambiguous adverse effect on their labor supply. In Study 2, in the payment scheme with implicit gender discrimination we do not see an adverse labor supply effect on average. However, there we do find a negative correlation between the belief in the presence of gender discrimination and labor supply, especially for women. This indicates that if women believe that they are discriminated against, even subtle gender discrimination can exert a negative effect on their labor supply. Given that many women believe that gender discrimination exists in the workplace and are also personally confronted with it (Parker and Funk, 2017; Investors in People, 2018), our results strongly suggest real adverse effects on women's labor supply due to gender discrimination in society.

IX. Conclusion

We provide the first causal evidence regarding how unfair chances to receive a low or high wage—stemming from an unspecified source or from gender discrimination—affect labor supply decisions. In our main study, we find that, at a given wage, explicit gender discrimination in chances considerably reduces the labor supply of disadvantaged workers compared to an equal-wage setting (-22%). This is the case even though workers only hurt themselves by working less. Moreover, low wages stemming from gender-discriminatory chances reduce labor supply almost twice as much as low wages resulting from fair chances (-13%). Interestingly, in the absence of discrimination, low-wage workers are insensitive to whether unequal wages result from fair or unfair chances. Advantaged workers are unresponsive to any type of inequality that we examine. Moreover, exploratory analysis suggests that men reduce their labor supply when they are in a disadvantaged wage position, whereas women decrease their labor supply only when their low wage can be attributed to gender-based discrimination in chances.

In a second study, we corroborate the existence of a negative effect of gender discrimination on labor supply, using discrimination in wages rather than in chances. We also document that workers' adverse labor supply reactions are proportional to their belief that they face negative discrimination. Furthermore, we find limited evidence that a positive gender wage gap in favor of women increases their labor supply—although outright positive discrimination does not—and we provide some evidence that positive discrimination of men lowers their labor supply.

Our findings provide a novel supply-side effect of gender discrimination in labor markets, and offer a complementary way to account for the lower supply of women and the gender gap in earnings. More broadly, our study opens new avenues for research on the reactions of workers who face discrimination.

References

Abeler, J., Altmann, S., Kube, S., and Wibral, M. (2010). Gift exchange and workers' fairness concerns: when equality is unfair. *Journal of the European Economic Association*, 8(6):1299–1324.

Adams, J. S. (1965). Inequity in social exchange. In Berkowitz, L., editor, *Advances in Experimental Social Psychology*, volume 2, pages 267–299. Academic Press, New York and London.

Akerlof, G. A. and Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics*, 105(2):255–283.

Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. In Ashenfelter, O. C. and Card, D., editors, *Handbook of Labor Economics*, volume 3c, chapter 48, pages 3143–3259. North Holland, Amsterdam.

Andreoni, J., Aydin, D., Barton, B., Bernheim, B. D., and Naecker, J. (2016). When fair isn't fair: sophisticated time inconsistency in social preferences. *Working Paper*.

Arechar, A. A., Gächter, S., and Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1):99–131.

Arrow, K. (1973). The theory of discrimination. In Ashenfelter, O. and Rees, A., editors, *Discrimination in Labor Markets*, pages 3–33. Princeton University Press, Princeton.

Balafoutas, L. and Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, 335(6068):579–582.

Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–75.

Becker, G. S. (1971). *The Economics of Discrimination*. University of Chicago Press, Chicago.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, pages 289–300.

Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95(2):94–98.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. In Duflo, E. and Banerjee, A., editors, *Handbook of Economic Field Experiments*, volume 1, chapter 8, pages 309–393. North Holland, Amsterdam.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–255.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *Quarterly Journal of Economics*, 130(2):571–614.

Bewley, T. (1999). *Why Don't Wages Fall in a Recession*. Harvard University Press, Cambridge.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Bohren, J. A., Imas, A., and Rosenberg, M. (2018). The dynamics of discrimination: Theory and evidence. *PIER Working Paper (18-016)*.

Bolton, G. E., Brandts, J., and Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *Economic Journal*, 115(506):1054–1076.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *Quarterly Journal of Economics*, 131(4):1753–1794.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–73.

Bracha, A., Gneezy, U., and Loewenstein, G. (2015). Relative pay and labor supply. *Journal of Labor Economics*, 33(2):297–315.

Breza, E., Kaur, S., and Shamdasani, Y. (2018). The morale effects of pay inequality. *Quarterly Journal of Economics*, 133(2):611–663.

Brock, J. M., Lange, A., and Ozbay, E. Y. (2013). Dictating the risk: Experimental evidence on giving in risky environments. *American Economic Review*, 103(1):415–37.

Bursztyn, L., Fujiwara, T., and Pallais, A. (2017). 'Acting wife': marriage market incentives and labor market investments. *American Economic Review*, 107(11):3288–3319.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics*, 129(3):1409–1447.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280):1433–1436.

Cappelen, A. W., Konow, J., Sørensen, E. Ø., and Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, 103(4):1398–1413.

Card, D., Cardoso, A. R., and Kline, P. (2015). Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*, 131(2):633–686.

Card, D., Mas, A., Moretti, E., and Saez, E. (2012). Inequality at work: the effect of peer salaries on job satisfaction. *American Economic Review*, 102(6):2981–3003.

Cettolin, E. and Riedl, A. (2016). Justice under uncertainty. *Management Science*, 63(11):3739–3759.

Charness, G. and Kuhn, P. (2007). Does pay inequality affect worker effort? Experimental evidence. *Journal of Labor Economics*, 25(4):693–723.

Chen, Y. and Konstan, J. (2015). Online field experiments: a selective survey of methods. *Journal of the Economic Science Association*, 1(1):29–42.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *Quarterly Journal of Economics*, 129(4):1625–1660.

Coffman, K. B., Coffman, L. C., and Ericson, K. M. M. (2016). The size of the LGBT population and the magnitude of antigay sentiment are substantially underestimated. *Management Science*, 63(10):3168–3186.

Coffman, K. B., Exley, C. L., and Niederle, M. (2017). When gender discrimination is not about gender. *Harvard Business School Working Paper*.

Cohn, A., Fehr, E., Herrmann, B., and Schneider, F. (2014). Social comparison and effort provision: evidence from a field experiment. *Journal of the European Economic Association*, 12(4):877–898.

Cortés, P. and Pan, J. (2019). When time binds: substitutes for household production, returns to working long hours, and the skilled gender wage gap. *Journal of Labor Economics*, 37(2):351–398.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

De Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302.

Diamond, P. A. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: comment. *Journal of Political Economy*, 75(5):765.

Dube, A., Giuliano, L., and Leonard, J. (2019). Fairness and frictions: the impact of unequal raises on quit behavior. *American Economic Review*, 109(2):620–63.

Dube, A., Jacobs, J., Naidu, S., and Suri, S. (2018). Monopsony in online labor markets. *NBER Working Paper (w24416)*.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252.

Epstein, L. G. and Segal, U. (1992). Quadratic social welfare functions. *Journal of Political Economy*, 100(4):691–712.

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

Flabbi, L. and Moro, A. (2012). The effect of job flexibility on female labor market outcomes: estimates from a search and bargaining model. *Journal of Econometrics*, 168(1):81–95.

Fleurbaey, M. (2010). Assessing risky social situations. *Journal of Political Economy*, 118:649–680.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In Ashenfelter, O. C. and Card, D., editors, *Handbook of Labor Economics*, volume 4a, chapter 1, pages 1–102. North Holland, Amsterdam.

Fréchette, G. R. (2012). Session-effects in the laboratory. *Experimental Economics*, 15(3):485–498.

Gächter, S. and Thöni, C. (2010). Social comparison and performance: experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behavior & Organization*, 76(3):531–543.

Ghani, E., Kerr, W. R., and Stanton, C. (2014). Diasporas and outsourcing: evidence from odesk and india. *Management Science*, 60(7):1677–1697.

Gilchrist, D. S., Luca, M., and Malhotra, D. (2016). When $3 + 1 > 4$: gift structure and reciprocity in the field. *Management Science*, 62(9):2639–2650.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: evidence from french grocery stores. *Quarterly Journal of Economics*, 132(3):1219–1260.

Goldin, C. (2014). A grand gender convergence: its last chapter. *American Economic Review*, 104(4):1091–1119.

Goldin, C. and Katz, L. F. (2016). A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics*, 34(3):705–746.

Goldin, C. and Rouse, C. (2000). Orchestrating impartiality: the impact of “blind” auditions on female musicians. *American Economic Review*, 90(4):715–741.

Grimalda, G., Kar, A., and Proto, E. (2016). Procedural fairness in lotteries assigning initial roles in a dynamic setting. *Experimental Economics*, 19(4):819–841.

Hammond, P. J. (1981). Ex-ante and ex-post welfare optimality under uncertainty. *Economica*, 48(191):235–250.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Horton, J. J. (2017a). The effects of algorithmic labor market recommendations: evidence from a field experiment. *Journal of Labor Economics*, 35(2):345–385.

Horton, J. J. (2017b). Price floors and employer preferences: Evidence from a minimum wage experiment. *CESifo Working Paper (6548)*.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425.

Ibañez, M. and Riener, G. (2018). Sorting through affirmative action: three field experiments in Colombia. *Journal of Labor Economics*, 36(2):437–478.

Investors in People (2018). Perceptions of gender discrimination in the workplace 2018. *Investors in People*.

Karni, E. and Safra, Z. (2002). Individual sense of justice: a utility representation. *Econometrica*, 70(1):263–284.

Karni, E., Salmon, T., and Sopher, B. (2008). Individual sense of fairness: an experimental study. *Experimental Economics*, 11(2):174–189.

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

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

Krawczyk, M. (2011). A model of procedural and distributive fairness. *Theory and Decision*, 70(1):111–128.

Krawczyk, M. and Le Lec, F. (2010). ’Give me a chance!’ An experiment in social decision under risk. *Experimental Economics*, 13(4):500–511.

Kuziemko, I., Norton, M. I., Saez, E., and Stantcheva, S. (2015). How elastic are preferences for redistribution? Evidence from randomized survey experiments. *American Economic Review*, 105(4):1478–1508.

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90:375 – 402. Social identity and discrimination.

Lang, K. (1986). A language theory of discrimination. *Quarterly Journal of Economics*, 101(2):363–382.

Leventhal, G. S., Karuza, J., and Fry, W. R. (1980). Beyond fairness: a theory of allocation preferences. *Justice and Social Interaction*, 3(1):167–218.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59.

Miao, B. and Zhong, S. (2018). Probabilistic social preference: how Machina's Mom randomizes her choice. *Economic Theory*, 65(1):1–24.

Mulligan, C. B. and Rubinstein, Y. (2008). Selection, investment, and women's relative wages over time. *Quarterly Journal of Economics*, 123(3):1061–1110.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3):799–866.

Neumark, D., Bank, R. J., and Van Nort, K. D. (1996). Sex discrimination in restaurant hiring: an audit study. *Quarterly Journal of Economics*, 111(3):915–941.

Niederle, M. (2016). Gender. In Kagel, J. H. and Roth, A. E., editors, *Handbook of Experimental Economics*, volume 2, chapter 8, pages 481–562. Princeton University Press, Princeton and Oxford.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1):1–16.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Ockenfels, A., Sliwka, D., and Werner, P. (2014). Bonus payments and reference point violations. *Management Science*, 61(7):1496–1513.

Pallais, A. (2014). Inefficient hiring in entry-level labor markets. *American Economic Review*, 104(11):3565–99.

Pallais, A. and Sands, E. G. (2016). Why the referential treatment? Evidence from field experiments on referrals. *Journal of Political Economy*, 124(6):1793–1828.

Parker, K. and Funk, C. (2017). Gender discrimination comes in many forms for todays working women. *Pew Research Center Fact Tank*.

Parsons, C. A., Sulaeman, J., Yates, M. C., and Hamermesh, D. S. (2011). Strike three: discrimination, incentives, and evaluation. *American Economic Review*, 101(4):1410–35.

Pascoe, E. A. and Smart Richman, L. (2009). Perceived discrimination and health: a meta-analytic review. *Psychological Bulletin*, 135(4):531.

Pfeffer, J. and Langton, N. (1993). The effect of wage dispersion on satisfaction, productivity, and working collaboratively: evidence from college and university faculty. *Administrative Science Quarterly*, 38(3):382–407.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62(4):659–661.

Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair womens careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408.

Riach, P. A. and Rich, J. (2002). Field experiments of discrimination in the market place. *Economic Journal*, 112(483):F480–F518.

Saito, K. (2013). Social preferences under risk: Equality of opportunity versus equality of outcome. *American Economic Review*, 103(7):3084–3101.

Sarsons, H., Gërxhani, K., Reuben, E., and Schram, A. (2019). Gender differences in recognition for group work. *Working Paper*.

Skarlicki, D. P. and Folger, R. (1997). Retaliation in the workplace: the roles of distributive, procedural, and interactional justice. *Journal of Applied Psychology*, 82(3):434.

Snowberg, E. and Yariv, L. (2018). Testing the waters: Behavior across participant pools. *NBER Working Paper (24781)*.

Stanton, C. T. and Thomas, C. (2015). Landing the first job: the value of intermediaries in online hiring. *Review of Economic Studies*, 83(2):810–854.

Trautmann, S. T. (2009). A tractable model of process fairness under risk. *Journal of Economic Psychology*, 30(5):803–813.

Trautmann, S. T. and van de Kuilen, G. (2016). Process fairness, outcome fairness, and dynamic consistency: Experimental evidence for risk and ambiguity. *Journal of Risk and Uncertainty*, 53(2-3):75–88.

Wiswall, M. and Zafar, B. (2017). Preference for the workplace, investment in human capital, and gender. *Quarterly Journal of Economics*, 133(1):457–507.

***** For Online Publication *****

Appendix

We provide additional figures and tables in Section A, examples screens of payment schemes in Section B, and the instructions of the experiment as seen by the workers in Section C. Each figure and table is referred to in the main text by its number.

A. Additional Figures and Tables

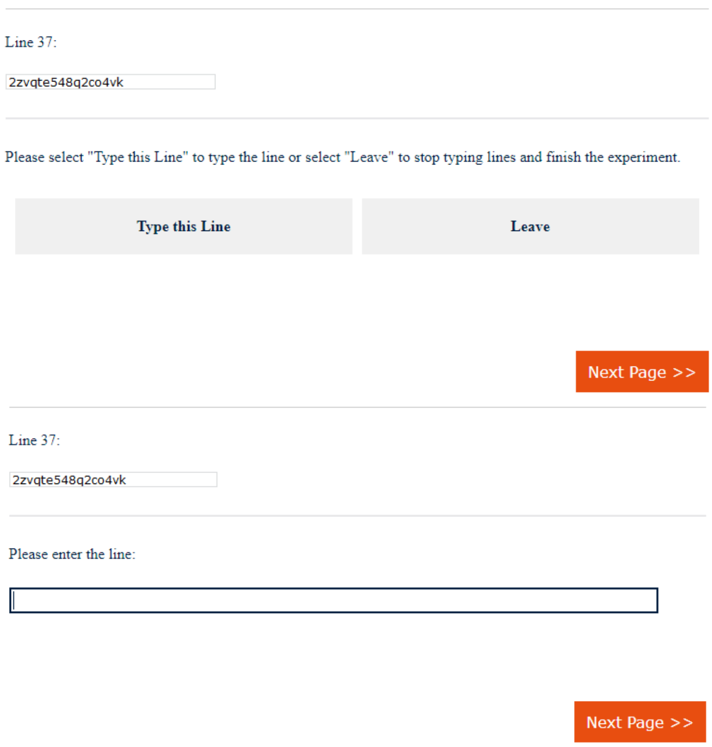


Figure A1: Task

Note: In the upper screenshot, the worker sees a line of characters, and decides whether to type the line or to leave the experiment. In the lower screenshot, if choosing not to leave, the worker is required to enter the line.

Study 1—Additional Information

Table A1: Study 1—Demographic Characteristics of Workers in Study 1

Demographic Characteristic	Mean (SD) or Percentage						F-test
	All Schemes	EQLOW	EQHIGH	UNEQ FAIR	UNEQ UNFAIR	UNEQ DISCR	(p-value)
Age	38 (12)	37 (13)	38 (13)	38 (12)	38 (13)	38 (12)	0.296
Task Experience on Platform	141 (176)	137 (154)	135 (157)	134 (151)	138 (176)	155 (219)	0.508
Woman	50%	49%	50%	51%	51%	49%	0.998
Student	16%	20%	16%	13%	15%	14%	0.177
UK National	93%	92%	94%	95%	93%	93%	0.608
Caucasian/White	88%	86%	91%	86%	89%	89%	0.748
Employed Full-Time	50%	44%	51%	52%	51%	54%	0.183
Employed Part-Time	20%	25%	22%	21%	18%	16%	0.101
Job Seeker	6%	5%	5%	7%	8%	5%	0.599
Not in Paid Work	18%	17%	19%	17%	18%	21%	0.813
Other Work Situation	5%	9%	4%	3%	5%	4%	0.580
N	1263–1271	127–128	128	255–257	292–294	294–296	

Note: N varies by characteristic because we could not obtain some characteristics from the platform for a few workers. F-test compares characteristics across the five payment schemes.

Table A2: Study 1—Labor Supply per Payment Scheme

	Low-Wage Workers					High-Wage Workers				
	Mean	SD	Min.	Max.	N	Mean	SD	Min.	Max.	N
EQLOW/EQHIGH	43.20	27.63	.04	.16	128	44.03	28.93	.04	.20	128
UNEQFAIR	37.44	29.16	.06	.17	125	43.23	29.64	.03	.20	127
UNEQUNFAIR	39.22	27.76	.04	.15	143	43.50	28.85	.02	.20	143
UNEQDISCR	33.62	26.41	.08	.11	143	45.74	26.72	.02	.16	145

Note: Labor supply is measured by the number of lines completed and ranges from 0 to 85. Min. and Max. indicate the percentage of workers completing the minimum and maximum number of lines.

Table A3: Study 1—*P*-Values of Six Predicted Differences in Labor Supply between Payment Schemes

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EqLow/EqHigh > UneqFair	0.017	0.170	0.050	0.212	0.371	0.387	1.000	1.000
EqLow/EqHigh > UneqUnfair	0.058	0.165	0.087	0.248	0.426	0.357	1.000	1.000
EqLow/EqHigh > UneqDiscr	0.001	0.001	0.006	0.002	0.853	0.607	1.000	0.910
UneqFair > UneqUnfair	0.878	0.513	0.878	0.513	0.780	0.473	1.000	0.946
UneqFair > UneqDiscr	0.081	0.013	0.102	0.025	0.905	0.720	0.905	0.900
UneqUnfair > UneqDiscr	0.021	0.006	0.031	0.019	0.886	0.752	1.000	0.752
N	539	533	539	533	543	542	543	542

Note: One-sided *p*-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing.

Table A4: Study 1—Tobit Regression of Labor Supply on Payment Schemes with Both Types of Workers

Scheme	All Workers
UNEQFAIR	-4.636
	(4.385)
UNEQUNFAIR	-4.432
	(4.098)
UNEQDISCR	-13.133^{***}
	(4.058)
UNEQFAIR × HighWage	2.030
	(6.318)
UNEQUNFAIR × HighWage	4.066
	(6.050)
UNEQDISCR × HighWage	13.307^{**}
	(5.790)
HighWage	2.291
	(4.352)
Controls	Yes
Prob $> F$	0.000
Pseudo R^2	0.008
N	1075

Restriction I

UNEQFAIR × HighWage = 0

UNEQUNFAIR × HighWage = UNEQFAIR × HighWage

UNEQDISCR × HighWage = UNEQUNFAIR × HighWage

Wald Test (two-sided p-value) = 0.073

Note: EQLOW serves as baseline. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$. The table presents the coefficient estimates. The overall null hypothesis, that there is no difference in labor supply between low-wage and high-wage workers in all payment scheme comparisons simultaneously, is represented by Restriction I at the bottom of the table. A Wald test marginally rejects this restriction (two-sided p-value = 0.073). Without controls, the p-value is 0.154.

Table A5: Study 1—Labor Supply per Gender and Payment Scheme

	Low-Wage Workers					High-Wage Workers				
	Mean	SD	Min.	Max.	N	Mean	SD	Min.	Max.	N
Men										
EQLOW/EQHIGH	44.40	27.20	.02	.18	65	42.83	27.95	.06	.16	64
UNEQFAIR	32.46	28.70	.10	.13	63	44.85	29.35	.03	.19	62
UNEQUNFAIR	36.13	27.44	.04	.14	72	45.13	30.05	.01	.23	69
UNEQDISCR	36.92	29.27	.06	.15	71	39.12	25.08	.04	.09	75
Women										
EQLOW/EQHIGH	41.95	28.24	.06	.14	63	45.23	30.05	.02	.25	64
UNEQFAIR	42.50	28.97	.03	.21	62	41.68	30.05	.03	.20	65
UNEQUNFAIR	42.37	27.92	.04	.15	71	41.97	27.80	.03	.16	74
UNEQDISCR	30.38	23.00	.11	.06	72	52.83	26.77	.03	.23	70

Note: Labor supply is measured by the number of lines completed and ranges from 0 to 85. Min. and Max. indicate the percentage of workers completing the minimum and maximum number of lines.

Table A6: Study 1—Tobit Regression of Labor Supply on Payment Schemes, for Men and Women

	Low-Wage Workers			High-Wage Workers		
	Men	Women	Men & Women	Men	Women	Men & Women
Scheme	(1)	(2)	(3)	(4)	(5)	(6)
UNEQFAIR	-12.788^{**}	3.919	-12.682^{**}	4.947	-5.478	3.238
	(5.950)	(6.311)	(5.909)	(6.307)	(6.674)	(6.351)
UNEQUNFAIR	-9.366^{*}	0.715	-8.896	5.069	-5.213	3.342
	(5.585)	(5.938)	(5.669)	(6.368)	(6.347)	(6.273)
UNEQDISCR	-10.802^{*}	-15.378^{***}	-11.092^{*}	-1.698	7.046	-3.679
	(5.952)	(5.561)	(5.831)	(5.640)	(6.390)	(5.649)
UNEQFAIR × Woman			17.242^{**}			-8.769
			(8.605)			(9.269)
UNEQUNFAIR × Woman			9.878			-9.287
			(8.163)			(8.848)
UNEQDISCR × Woman			-5.021			10.430
			(7.966)			(8.533)
Woman			-5.400			5.317
			(6.212)			(6.854)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Prob $> F$	0.053	0.000	0.000	0.045	0.000	0.003
Pseudo R^2	0.014	0.019	0.013	0.013	0.010	0.010
N	268	265	542	270	272	542

Note: EQLOW/EQHIGH serves as baseline. Standard errors are indicated in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A7: Study 1—Tests of Inequalities in Labor Supply between Payment Schemes, for Men and Women

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
Men								
EQLOW/EQHIGH > UNEQFAIR	0.002	0.016	0.006	0.049	0.818	0.783	0.818	0.783
UNEQFAIR > UNEQUNFAIR	0.920	0.716	0.920	0.716	0.248	0.508	0.372	0.762
UNEQUNFAIR > UNEQDISCR	0.238	0.398	0.358	0.598	0.063	0.123	0.190	0.369
N	271	268	271	268	270	270	270	270
Women								
EQLOW/EQHIGH > UNEQFAIR	0.767	0.732	0.767	0.732	0.104	0.206	0.311	0.619
UNEQFAIR > UNEQUNFAIR	0.751	0.305	1.000	0.457	0.801	0.517	1.000	0.775
UNEQUNFAIR > UNEQDISCR	0.003	0.001	0.009	0.004	0.994	0.981	0.994	0.981
N	268	265	268	265	273	272	273	272

Note: One-sided p-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing. See Table A8 for the six pairwise comparisons of schemes by gender.

Table A8: Study 1—*P*-Values of Six Predicted Differences in Labor Supply between Payment Schemes, for Men and Women

Men

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EQLOW/EQHIGH > UNEQFAIR	0.002	0.016	0.012	0.098	0.818	0.783	0.818	0.979
EQLOW/EQHIGH > UNEQUNFAIR	0.021	0.047	0.041	0.095	0.818	0.787	1.000	0.787
EQLOW/EQHIGH > UNEQDISCR	0.018	0.035	0.055	0.106	0.111	0.382	0.222	0.764
UNEQFAIR > UNEQUNFAIR	0.920	0.716	0.920	0.716	0.248	0.508	0.372	0.762
UNEQFAIR > UNEQDISCR	0.912	0.630	1.000	0.787	0.065	0.122	0.196	0.731
UNEQUNFAIR > UNEQDISCR	0.238	0.398	0.358	0.598	0.063	0.123	0.380	0.369
N	271	268	271	268	270	270	270	270

Women

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EQLOW/EQHIGH > UNEQFAIR	0.767	0.732	0.958	0.732	0.104	0.206	0.621	1.000
EQLOW/EQHIGH > UNEQUNFAIR	0.768	0.548	0.768	0.685	0.139	0.206	0.418	0.618
EQLOW/EQHIGH > UNEQDISCR	0.005	0.003	0.009	0.006	0.972	0.864	1.000	1.000
UNEQFAIR > UNEQUNFAIR	0.751	0.305	1.126	0.457	0.801	0.517	1.000	1.000
UNEQFAIR > UNEQDISCR	0.004	0.001	0.012	0.004	0.996	0.977	0.996	1.000
UNEQUNFAIR > UNEQDISCR	0.003	0.001	0.018	0.004	0.994	0.981	1.000	0.981
N	268	265	268	265	273	272	273	272

Note: One-sided *p*-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing.

Table A9: Study 1—*P*-Value of Test that Positive Discrimination of Women does not Affect their Labor Supply

	High-Wage Women	
Inequality	(1)	(2)
Technique	Dunn	Tobit
EQHIGH ≠ UNEQDISCR	0.055	0.271
UNEQFAIR ≠ UNEQDISCR	0.007	0.045
UNEQUNFAIR ≠ UNEQDISCR	0.012	0.038
N	273	272

Note: Two-sided *p*-values are presented. The Tobit estimates come from specification (5) of Table A6.

Table A10: Study 1—Hurdle Model (Labor Supply on Payment Schemes), for Low-Wage Workers

Scheme	Low-Wage Workers
Extensive Margin	
UNEQFAIR	-4.417
	(5.927)
UNEQUNFAIR	-4.406
	(5.413)
UNEQDISCR	-15.442^{***}
	(5.715)
Intensive Margin	
UNEQFAIR	-0.516^{*}
	(0.302)
UNEQUNFAIR	-0.304
	(0.316)
UNEQDISCR	-0.773^{**}
	(0.307)
Controls	Yes
Prob $> F$	0.032
Pseudo R^2	0.013
N	533

Note: EQLOW serves as baseline. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A11: Study 1—P-Values of Six Predicted Differences in Labor Supply between Payment Schemes, Hurdle Model

	Low-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)
Margin	Extensive		Intensive	
BH Correction	No	Yes	No	Yes
EQLOW/EQHIGH > UNEQFAIR	0.228	0.285	0.044	0.087
EQLOW/EQHIGH > UNEQUNFAIR	0.208	0.312	0.168	0.210
EQLOW/EQHIGH > UNEQDISCR	0.003	0.021	0.006	0.035
UNEQFAIR > UNEQUNFAIR	0.501	0.501	0.787	0.787
UNEQFAIR > UNEQDISCR	0.039	0.078	0.138	0.206
UNEQUNFAIR > UNEQDISCR	0.026	0.079	0.029	0.086
N	533	533	533	533

Note: One-sided p-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing.

Table A12: Study 1—Tobit Regression of Labor Supply on Payment Schemes, for Young (age ≤ 36) and Old (age > 36)

	Low-Wage Workers		High-Wage Workers	
	Young	Old	Young	Old
Scheme	(1)	(2)	(3)	(4)
UNEQFAIR	−5.633	−4.065	−9.100	5.953
	(6.245)	(6.183)	(7.093)	(6.122)
UNEQUNFAIR	−10.527*	0.962	−5.550	1.650
	(6.099)	(5.712)	(6.463)	(6.260)
UNEQDISCR	−18.440***	−9.274*	−0.442	3.300
	(6.241)	(5.360)	(6.158)	(5.940)
Controls	Yes	Yes	Yes	Yes
Prob $> F$	0.034	0.088	0.000	0.625
Pseudo R^2	0.013	0.014	0.011	0.007
N	260	273	290	252

Note: Median age is 36 in the sample. EQLOW/EQHIGH serves as baseline. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A13: Study 1—*P*-Values of Six Predicted Differences in Labor Supply between Payment Schemes, for Young (age ≤ 36) and Old (age > 36)

Young

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EQLOW/EQHIGH > UNEQFAIR	0.086	0.184	0.108	0.230	0.054	0.100	0.321	0.602
EQLOW/EQHIGH > UNEQUNFAIR	0.047	0.043	0.094	0.086	0.106	0.196	0.318	0.587
EQLOW/EQHIGH > UNEQDISCR	0.003	0.002	0.018	0.010	0.751	0.471	1.000	0.943
UNEQFAIR > UNEQUNFAIR	0.187	0.207	0.187	0.207	0.851	0.712	1.000	1.000
UNEQFAIR > UNEQDISCR	0.036	0.017	0.108	0.052	0.952	0.918	0.952	0.918
UNEQUNFAIR > UNEQDISCR	0.058	0.088	0.087	0.132	0.901	0.823	1.000	1.000
N	263	260	263	260	291	290	291	290

Old

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EQLOW/EQHIGH > UNEQFAIR	0.032	0.256	0.096	0.384	0.877	0.834	0.877	0.834
EQLOW/EQHIGH > UNEQUNFAIR	0.191	0.567	0.287	0.709	0.791	0.604	1.000	1.000
EQLOW/EQHIGH > UNEQDISCR	0.032	0.042	0.190	0.127	0.863	0.711	1.000	0.888
UNEQFAIR > UNEQUNFAIR	0.950	0.792	0.950	0.792	0.162	0.244	0.974	1.000
UNEQFAIR > UNEQDISCR	0.755	0.186	0.944	0.371	0.229	0.322	0.687	0.966
UNEQUNFAIR > UNEQDISCR	0.049	0.023	0.099	0.140	0.821	0.607	1.000	0.911
N	276	273	276	273	252	252	252	252

Note: One-sided *p*-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing.

Table A14: Study 1—Tobit Regression of Labor Supply on Payment Schemes, for Employed Full Time and Not Employed Full Time

	Low-Wage Workers		High-Wage Workers	
	Full-Time	Not Full-Time	Full-Time	Not Full-Time
Scheme	(1)	(2)	(3)	(4)
UNEQFAIR	−7.091	0.891	−1.010	−3.379
	(6.546)	(5.955)	(6.384)	(6.724)
UNEQUNFAIR	−7.809	−0.410	2.169	−5.549
	(6.109)	(5.661)	(6.374)	(6.327)
UNEQDISCR	-10.922^{*}	-15.172^{***}	5.921	−5.965
	(6.074)	(5.578)	(5.872)	(6.239)
Controls	Yes	Yes	Yes	Yes
Prob > F	0.301	0.002	0.003	< 0.001
Pseudo R^2	0.009	0.017	0.004	0.017
N	254	279	296	246

Note: Workers who are full-time employed account for 51% of the sample. Those who are not full-time employed are: part-time employed (20%), unemployed (6%), not in a paid job (18%) or other (5%). EQLOW/EQHIGH serves as baseline. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A15: Study 1—*P*-Values of Six Predicted Differences in Labor Supply between Payment Schemes, for Employed Full Time and Not Employed Full Time

Employed Full Time

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EqLow/EqHigh > UneqFair	0.033	0.140	0.066	0.280	0.183	0.437	1.000	1.000
EqLow/EqHigh > UneqUnfair	0.029	0.101	0.086	0.304	0.752	0.633	1.000	1.000
EqLow/EqHigh > UneqDiscr	0.016	0.037	0.094	0.220	0.944	0.843	1.000	1.000
UneqFair > UneqUnfair	0.244	0.455	0.244	0.455	0.821	0.694	1.000	1.000
UneqFair > UneqDiscr	0.183	0.256	0.275	0.384	0.973	0.889	0.973	0.889
UneqUnfair > UneqDiscr	0.187	0.280	0.234	0.349	0.947	0.749	1.000	1.000
N	254	254	254	254	296	296	296	296

Not Employed Full Time

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes	No	No	Yes	Yes
EqLow/EqHigh > UneqFair	0.089	0.559	0.133	0.559	0.249	0.308	0.249	0.616
EqLow/EqHigh > UneqUnfair	0.216	0.471	0.270	0.589	0.221	0.191	0.286	0.572
EqLow/EqHigh > UneqDiscr	0.004	0.004	0.023	0.021	0.165	0.170	0.990	1.000
UneqFair > UneqUnfair	0.889	0.415	0.889	0.622	0.222	0.361	0.278	0.452
UneqFair > UneqDiscr	0.041	0.005	0.082	0.009	0.167	0.338	0.502	0.507
UneqUnfair > UneqDiscr	0.006	0.004	0.018	0.011	0.190	0.471	0.380	0.471
N	279	279	279	279	246	246	246	246

Note: One-sided *p*-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing.

Table A16: Study 1—Tobit Regression of Time Spent in the Experiment on Payment Schemes

	Low-Wage Workers	High-Wage Workers
Scheme	(1)	(2)
UNEQFAIR	-1.478	-0.729
	(1.935)	(2.002)
UNEQUNFAIR	-0.720	-0.062
	(1.803)	(1.946)
UNEQDISCR	-4.621^{***}	2.060
	(1.760)	(1.923)
Controls	Yes	Yes
Prob $> F$	0.002	0.004
Pseudo R^2	0.009	0.008
N	533	542

Note: Time spent in the experiment is measured in minutes. Mean times are 22.65 (SD = 14.96) minutes for low-wage workers, and 28.19 (SD = 16.02) minutes for high-wage workers. Note that our time measure is not necessarily time worked since it starts when workers begin reading the instructions and ends when they quit the experiment or at 65 minutes if they have not quit by then. Moreover, this measure does not account for the breaks that workers can take. Therefore, it is most likely not an appropriate measure of labor supply. EQLOW/EQHIGH serves as baseline. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A17: Study 1—P-Values of Six Predicted Differences in Time Spent in the Experiment between Payment Schemes

	Low-Wage Workers				High-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit	Dunn	Tobit
BH Correction	No	Yes	No	Yes	No	Yes	No	Yes
EQLOW/EQHIGH > UNEQFAIR	0.028	0.223	0.057	0.334	0.148	0.358	0.889	1.000
EQLOW/EQHIGH > UNEQUNFAIR	0.180	0.345	0.225	0.431	0.247	0.487	0.741	1.000
EQLOW/EQHIGH > UNEQDISCR	0.006	0.004	0.037	0.027	0.921	0.858	1.000	1.000
UNEQFAIR > UNEQUNFAIR	0.949	0.657	0.949	0.657	0.852	0.639	1.000	1.000
UNEQFAIR > UNEQDISCR	0.137	0.042	0.205	0.083	0.907	0.929	0.970	0.929
UNEQUNFAIR > UNEQDISCR	0.013	0.010	0.039	0.029	0.926	0.879	1.000	1.000
N	539	533	539	533	543	542	543	542

Note: One-sided p-values are presented, in the direction predicted. BH corrections account for multiple hypothesis testing.

Table A18: Study 1—Labor Supply per Payment Scheme, for Workers not Beating the Odds and Workers Beating the Odds

Low-Wage Workers	Mean	SD	Min.	Max.	*N*
Workers *not* Beating the Odds					
EqLow	43.20	27.63	.04	.16	128
UneqFair	37.44	29.16	.06	.17	125
UneqUnfair	39.22	27.76	.04	.15	143
UneqDiscr	33.62	26.41	.08	.11	143
Workers Beating the Odds					
UneqUnfair(BeatOdds)	37.98	27.58	.02	.16	50
UneqDiscr(BeatOdds)	38.15	32.10	.04	.24	46
High-Wage Workers	**Mean**	**SD**	**Min.**	**Max.**	***N***
Workers *not* Beating the Odds					
EqHigh	44.03	28.93	.04	.20	128
UneqFair	43.23	29.64	.03	.20	127
UneqUnfair	43.50	28.85	.02	.20	143
UneqDiscr	45.74	26.72	.02	.16	145
Workers Beating the Odds					
UneqUnfair(BeatOdds)	42.55	28.34	.02	.17	47
UneqDiscr(BeatOdds)	40.72	27.79	.02	.15	46

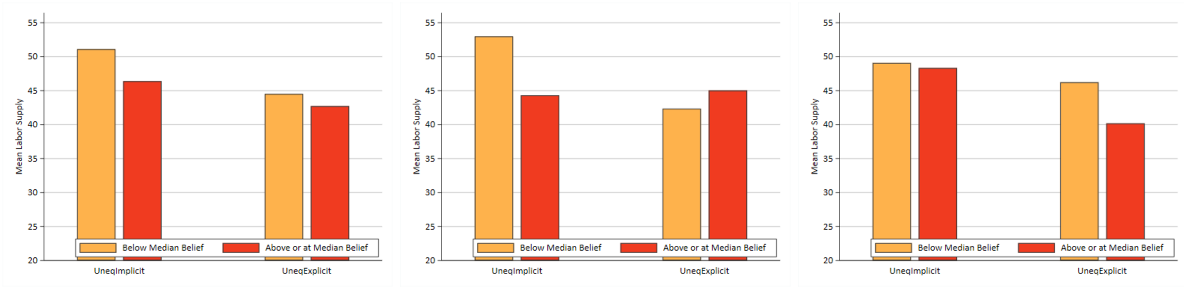
Note: Labor supply is measured by the number of lines completed and ranges from 0 to 85. Min. and Max. indicate the percentage of workers completing the minimum and maximum number of lines. Low-wage workers beating the odds had high chances, and high-wage workers beating the odds had low chances.

Table A19: Study 1—*P*-Values of Tests on Difference in Labor Supply between Payment Schemes, for Workers Beating the Odds

	Low-Wage Workers		High-Wage Workers	
Predicted Inequality	(1)	(2)	(3)	(4)
Technique	Dunn	Tobit	Dunn	Tobit
UneqFair > UneqUnfair(BeatOdds)	0.181	0.316	0.241	0.454
UneqUnfair(BeatOdds) > UneqDiscr(BeatOdds)	0.811	0.658	0.192	0.279
N	221	219	220	219

Notes: One-sided *p*-values are presented, in the direction predicted by the model.

Study 2—Additional Information



(a) All Workers (b) Women (c) Men

Figure A2: Study 2—Mean Labor Supply per Payment Scheme Split by Median Belief in that Gender Discrimination was Used in the Experiment
Note: Labor supply is measured by the number of lines completed and ranges from 0 to 85. Workers answered the post-experiment question "*During the task, did you believe that gender discrimination was used to determine your payment per line?*" on a Likert scale from 1 (*Not at all*) to 7 (*Completely*).

Table A20: Study 2—Demographic Characteristics of Workers

Demographic Characteristic	Mean (SD) or Percentage				F-test
	All Schemes	Uneq	UneqIm	UneqEx	(p-value)
Age	35 (12)	35 (12)	35 (12)	35 (12)	0.972
Task Experience on Platform	101 (124)	92 (114)	105 (135)	104 (123)	0.119
Woman	50	50	50	50	0.999
Student	20	20	21	21	0.937
UK National	93	94	92	93	0.434
Employed Full-Time	51	51	51	51	0.987
Employed Part-Time	19	21	17	20	0.144
Job Seeker	9	6	10	9	0.215
Not in Paid Work	15	15	15	15	0.965
Other Work Situation	6	6	7	5	0.339
N	1576–1593	523–531	524–532	524–530	

Note: N varies by characteristic because we could not obtain some characteristics from the platform for a few workers.

Table A21: Study 2—Labor Supply per Payment Scheme

	Low-Wage Workers					High-Wage Workers				
	Mean	SD	Min.	Max.	N	Mean	SD	Min.	Max.	N
UNEQ	47.69	28.60	.04	.24	265	53.16	29.44	.02	.33	266
UNEQIM	48.99	29.61	.05	.27	266	53.12	29.19	.03	.35	266
UNEQEX	43.29	29.90	.06	.18	265	50.02	30.51	.06	.29	265

Note: Labor supply is measured by the number of lines completed and ranges from 0 to 85. Min. and Max. indicate the percentage of workers completing the minimum and maximum number of lines.

Table A22: Study 2—Labor Supply per Gender and Payment Scheme

	Low-Wage Workers					High-Wage Workers				
	Mean	SD	Min.	Max.	N	Mean	SD	Min.	Max.	N
Men										
UNEQ	46.73	29.74	.05	.27	133	50.53	29.61	.03	.32	133
UNEQIM	48.73	30.21	.03	.29	132	45.87	30.14	.05	.26	134
UNEQEX	42.28	29.63	.04	.16	133	44.93	30.33	.08	.22	132
Women										
UNEQ	48.66	27.48	.03	.21	132	55.80	29.14	.01	.35	133
UNEQIM	49.25	29.12	.07	.26	134	60.48	26.33	.01	.43	132
UNEQEX	44.30	30.25	.05	.20	132	55.08	29.96	.04	.36	133

Note: Labor supply is measured by the number of lines completed and ranges from 0 to 85. Min. and Max. indicate the percentage of workers completing the minimum and maximum number of lines.

Table A23: Study 2—Tobit Regressions of Labor Supply on Payment Schemes

	Low-Wage Workers	High-Wage Workers
Scheme	(1)	(2)
UNEQIM	1.066	−0.176
	(3.608)	(3.915)
UNEQEX	-7.131^{**}	−4.267
	(3.443)	(3.907)
Controls	Yes	Yes
Prob > F	0.074	0.000
Pseudo R^2	0.005	0.009
N	778	779

Note: UNEQ serves as baseline. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A24: Study 2—P-values of Three Predicted Differences in Labor Supply between Payment Schemes

	Low-Wage Workers			
Predicted Inequality	(1)	(2)	(3)	(4)
Technique	Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes
UNEQ > UNEQIM	0.843	0.616	0.843	0.616
UNEQ > UNEQEX	0.013	0.019	0.020	0.029
UNEQIM > UNEQEX	0.004	0.012	0.012	0.035
N	796	778	796	778

Note: One-sided p-values in the direction predicted. BH corrections account for multiple hypothesis testing. Tobit regressions results are presented in Table A23.

Table A25: Study 2—P-values of Kruskal-Wallis and Wald tests across all Payment Schemes, for Men and Women

	Low-Wage Workers		High-Wage Workers	
	KW test	Wald test	KW test	Wald test
Men				
UNEQ = UNEQIM = UNEQEX	0.122	0.136	0.231	0.199
N	398	391	399	388
Women				
UNEQ = UNEQIM = UNEQEX	0.277	0.260	0.205	0.653
N	398	387	398	391

Note: Kruskal-Wallis (KW) and Wald tests are both two-sided.

Table A26: Study 2—P-values of Three Predicted Differences in Labor Supply between Payment Schemes, for Men and Women

	Low-Wage Workers					High-Wage Workers			
Predicted (In)equality	(1)	(2)	(3)	(4)	Predicted Inequality	(5)	(6)	(7)	(8)
Technique	Dunn	Tobit	Dunn	Tobit		Dunn	Tobit	Dunn	Tobit
BH Correction	No	No	Yes	Yes		No	No	Yes	Yes
Women									
UNEQ > UNEQIM	0.783	0.450	0.783	0.450	UNEQ < UNEQIM	0.035	0.182	0.104	0.546
UNEQ > UNEQEX	0.049	0.066	0.073	0.198	UNEQ < UNEQEX	0.773	0.390	1.000	0.584
UNEQIM > UNEQEX	0.036	0.092	0.107	0.139	UNEQIM < UNEQEX	0.973	0.733	0.973	0.733
N	398	387	398	387		398	391	398	391
Men									
UNEQ = UNEQIM	0.300	0.660	0.300	0.660	UNEQ > UNEQIM	0.050	0.185	0.076	0.185
UNEQ = UNEQEX	0.072	0.151	0.109	0.227	UNEQ > UNEQEX	0.023	0.036	0.069	0.109
UNEQIM = UNEQEX	0.024	0.064	0.071	0.193	UNEQIM > UNEQEX	0.169	0.179	0.169	0.268
N	398	391	398	391		399	388	399	388

Note: One-sided p-values are presented in the direction predicted if we made a prediction; two-sided p-values are used for low-wage men. BH corrections account for multiple hypothesis testing.

Table A27: Study 2—Tobit Regressions of Labor Supply on Payment Schemes, for Men and Women

	Low-Wage Workers			High-Wage Workers		
	Men	Women	Men & Women	Men	Women	Men & Women
Scheme	(1)	(2)	(3)	(4)	(5)	(6)
UNEQIM	2.347	−0.636		−4.740	−0.636	
	(5.332)	(5.024)		(5.276)	(5.024)	
UNEQEX	−6.874	−7.518		−9.596*	−7.518	
	(4.779)	(4.976)		(5.331)	(4.976)	
UNEQIM/EX			−3.152			−7.524
			(4.443)			(4.670)
UNEQIM/EX × Woman			0.127			10.752
			(6.204)			(6.661)
Woman			1.055			5.865
			(5.167)			(5.504)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Prob > F	0.046	0.398	0.299	0.002	0.073	0.000
Pseudo R^2	0.009	0.006	0.004	0.013	0.011	0.010
N	391	387	778	388	391	779

Note: UNEQ serves as baseline. Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A28: Study 2—Tobit Regressions of Low-Wage Workers' Labor Supply on Belief that Discrimination was Used in Experiment

	All Schemes				UNEQIM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Belief Discrimination was Used	−1.246**	−1.576***	−1.740***	−1.774***	−1.502	−2.075*	−2.492**	−2.332*
	(0.594)	(0.609)	(0.605)	(0.601)	(1.169)	(1.245)	(1.246)	(1.252)
Belief Paired Worker Exists			1.401***	1.401*			3.061**	1.954
			(0.756)	(0.807)			(1.380)	(1.571)
Trust Instructions were Truthful				3.474***				3.499
				(1.100)				(2.211)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Prob > F	0.036	0.077	0.003	0.001	0.200	0.565	0.180	0.106
Pseudo R^2	0.001	0.005	0.007	0.008	0.001	0.009	0.012	0.013
N	788	771	771	771	264	260	260	260

Note: The variable *Belief Discrimination was Used* ranges from 1 (*Not at all)* to 7 (*Completely*). We employ several Tobit regression specifications that vary the inclusion of controls and two variables that serve as proxy for how much participants believe that the information provided in the experiment is real (*Belief Paired Worker Exists* and *Trust Instructions were Truthful*). Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table A29: Study 2—Tobit Regressions of Low-Wage Workers' Labor Supply on Belief that Discrimination was Used in Experiment in UNEQIM, for Men and Women

	Men				Women			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Belief Discrimination was Used	0.287	1.079	0.910	0.990	-3.154^{*}	-4.223^{**}	-4.750^{***}	-4.407^{***}
	(1.654)	(1.575)	(1.576)	(1.607)	(1.633)	(1.728)	(1.708)	(1.651)
Belief Paired Worker Exists			1.419	1.158			3.813^{**}	0.292
			(2.092)	(3.406)			(1.823)	(2.069)
Trust Instructions were Truthful				1.110				9.143^{***}
				(1.962)				(2.798)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Prob $> F$	0.863	0.002	0.002	0.002	0.056	0.101	0.028	< 0.001
Pseudo R^2	0.000	0.029	0.029	0.030	0.004	0.025	0.030	0.042
N	132	130	130	130	132	130	130	130

Note: The variable *Belief Discrimination was Used* ranges from 1 (*Not at all)* to 7 (*Completely*). We employ several Tobit regression specifications that vary the inclusion of controls and two variables that serve as proxy for how much participants believe that the information provided in the experiment is real (*Belief Paired Worker Exists* and *Trust Instructions were Truthful*). Standard errors are in parentheses. Two-sided p-values: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

B. Example Screens of Payment Schemes

Payment per Line

In this part you can earn a payment with each line you correctly enter. You will now be informed about the amount of this payment per line.

We also inform you about the payment per line of another participant in this economic experiment. This other participant is a real person, participates in the exact same experiment as you do, and has been recruited from the same on-line work platform as you. The other participant also finished the Practice Part and faces the exact same task as you. This means that the other participant faces the exact same lines as you in the task.

Your payment and the payment of the other participant are **decided through a computerized lottery**. One of you will receive a payment per line of £0.06, the other will receive a payment per line of £0.03.

You are a woman, therefore you have:
a **25% chance** of receiving the payment per line of **£0.06** and
a **75% chance** of receiving the payment per line of **£0.03**.

The other participant is a man, therefore he has:
a **75% chance** of receiving the payment per line of **£0.06**, and
a **25% chance** of receiving the payment per line of **£0.03**.

The software will inform you about the outcome of the lottery on the next page.

Once you have learned your payment per line, you can start the task.

(a) Presentation of the Payment Scheme (UNEQDISCR in Study 1)

The lottery outcome is the following:

Your payment per line is **£0.03**.
The other participant's payment per line is **£0.06**.

(b) Presentation of the Wages (when own wage is low in UNEQDISCR in Study 1)

Figure A3: Presentation of the Payment Scheme and Presentation of Wages

C. Instructions of the Experiment

Invitation Email

[Note: Potential participants see the time limit that they have to complete the experiment]

We would like to invite you to participate in an online economic experiment about decision making. You will be paid a reward of 0.70 for about 5 minutes of participation. Thereafter, as will be explained in the instructions, you can earn more money with the decisions you make by participating in this experiment for a longer time.

IMPORTANT: All information provided will be collected and stored ANONYMOUSLY.

****

You receive this invitation because you are registered at Prolific. Please consult the Prolific website in case you want your data to be removed from the platform.

Instructions

[Notes: Participants are provided with a countdown from 65 to 0 minutes.]

Instructions

Welcome to this economic experiment,

You can earn a considerable amount of money with the decisions you make. Please read these instructions carefully. Importantly, unlike experiments in some other social sciences, economic experiments employ a strict non-deception policy. This means that all information you receive is truthful.

The only way to leave this economic experiment and be paid is to click on the button "Leave" and go to the next page. Once you do this, you will see a message that the experiment is now over and that you can close your browser page. You will not be paid if you leave at any moment by closing your browser window without clicking on the button "Leave" and going to the next page that tells you that the experiment is over.

This economic experiment consists of a Practice Part, where you cannot yet earn money, followed by a Task Part where you can earn money. The Practice Part consists of these Instructions, some comprehension questions, and a practice exercise. It is important that you answer the comprehension questions correctly by yourself. Please do not consult other people when answering these questions. In case you do not answer a question correctly, you will have two more chances to correct your answer. If you do not answer all questions correctly after these two additional chances, you will not be able to participate in the Task Part and the experiment ends for you. In that case you will be paid £0.45. When you have answered all comprehension questions correctly you can participate in the Task Part. In the Task Part, you can earn money by working on a task. You can stop working on the task whenever you prefer.

Recall, that to leave this economic experiment and to **be paid** you need to click on the "Leave " button and go to the next page. Once you have done this, you will see a message that the experiment is over for you and that you can close your browser page. You will **not be paid** if you leave the experiment without following the described procedure.

Note that you cannot leave the experiment and be paid before you finish the Practice Part (which lasts about 5 minutes). Thereafter you can leave the experiment at any time.

After you leave the experiment using the "Leave button, the money you have earned will be paid to you through Prolific.

Task Part

In the Task Part of this experiment, you can earn money by working on a task. You can decide how much of the task you want to complete. The task is to enter preset lines of random numbers and/or letters on your computer. You will receive a payment for each line you copy correctly. Nobody else than yourself will derive any earnings from your work, including the experimenters. The lines of numbers and/or letters you enter have no further use for anyone.

You will see one line at the time. Once you have entered a line correctly, you can go to the next page to see the next line. Each time you see a new line, you can decide whether you would like to type this line or leave the experiment.

In case you make a mistake when entering the line, the software will tell you so. You will need to correct this mistake before you can proceed to the next line.

The length of the sequences of random numbers and/or letters will increase as you complete more lines. That is, lines will be relatively short at the beginning but get longer over time.

You will be informed of your payment per line at the beginning of the Task Part.

In the Task Part you may also receive anonymous payment information regarding another participant.

Leave the Experiment

You can stop entering lines at any moment. Note, however, that the only way **to stop and to be paid** is to **click on the "Leave" button and then go to the next page**. You will then see a message that the experiment is over for you, that you need to click on a Prolific link to validate your participation, and that you can then close your browser window.

You will see the Leave button whenever you are presented a new line. If you decide to leave, you will not be able to start working again. That is, **once you leave the experiment you cannot go back**.

Payment

When you leave the experiment according to the described procedure you will receive a **payment per line you entered correctly**. You will be informed about the amount of the payment per line when you see the first line to be entered. In addition, you will also receive a **fixed amount of £0.70**, irrespective of the number of lines entered.

Decision

The decision you make in this economic experiment is to choose how much of the task you want to complete. You are the only one deciding how much you work. Your decision only affects your own earnings.

Practice Part

In the Practice Part, we ask you to correctly answer a number of comprehension questions. It is important that you answer these comprehension questions by yourself. For each question you will have three chances. If you do not correctly answer all comprehension questions you will not be able to participate in the Task Part. In this case the experiment will end for you and you will be paid 0.45.

After having correctly answered all comprehension questions, you will enter two practice lines to make you familiar with the task. Neither can you earn payments per line with these practice lines, nor will these practice lines affect the Task Part in any way.

This is the end of the instructions.

I confirm that I have read the instructions carefully and I am ready to start the Practice Part. I will not be able to go back to the instructions once I go to the next page.