Anticipated Discrimination, Choices, and Performance: Experimental Evidence*

February 17, 2020

GARY CHARNES1, RAMÓN COBO-REYES2, SIMONE MERAGLIA3, AND ÁNGELA SÁNCHEZ4

1 UNIVERSITY OF CALIFORNIA, SANTA BARBARA

2 AMERICAN UNIVERSITY OF SHARJAH

3 UNIVERSITY OF EXETER

4 NYU ABU DHABI

Abstract

This paper studies experimentally whether potential perceived discrimination affects decisions in a labor-market setting with different stereotypes. Participants are assigned to a seven-person group and randomly allocated a role as a firm or worker. In each group, there are five workers and two firms. The only information firms have about each worker is a self-selected avatar (male, female or neutral) representing a worker’s gender. Each firm then decides which worker to hire. Female workers react to potential discrimination when they know the task is math-related, but not otherwise. Men choose similar avatar patterns regardless of the task. Men do perform at much higher levels in the math-related task, but there is no difference in performance in the emotion-recognition task, where there is a strong female stereotype.

JEL numbers: A14, C91, J16, J70

Keywords: Gender stereotypes, discrimination, hiring patterns

* We are particularly indebted to Gabriel Katz and Brianna Halladay for very useful comments and discussions. We also thank seminar participants at the American University of Sharjah, University of Exeter BID Brown Bag seminar, and participants at the 2017 ESA Meeting in San Diego. The authors are thankful to the University of Exeter BID and BIP Clusters and the American University of Sharjah FRG Grant for financially supporting this project.
1. Introduction

A recent and growing literature points to expectations as an important predictor of educational choices, showing that, along with ability perceptions, beliefs about future earnings are one of the determinants of choices of college major (Arcidiacono et al., 2012; Zafar, 2013; Wiswall and Zafar, 2015). Among other factors affecting beliefs, stereotypes (Bordalo et al., 2016; Coffman, 2014) and related discrimination issues (Bohnet, van Geen, and Bazerman, 2016; Reuben, Sapienza, and Zingales, 2014) have been extensively analyzed in the literature.\(^1\)

While stereotypes and labor-market discrimination have been already documented, nevertheless little is known regarding whether and how people react when they perceive potential discrimination.\(^2\) This anticipation could be even more important than the discrimination itself for the agents’ decision process. Consider, for example, the case of science, technology, engineering, or mathematics (STEM) careers.\(^3\) A woman who, when planning what major to choose, anticipates that she will face discrimination (either a lower probability of being hired or a higher probability of getting a lower salary than her male counterpart) in a particular field might react strategically to this perceived discrimination by reducing the likelihood of choosing a degree that relates to some particular profession.

Different factors have been proposed to explain the low representation of women in science. First, gender differences in preferences (Croson and Gneezy, 2009; Niederle and

\(^1\) See also the extensive literature on statistical discrimination (Lundberg and Startz, 1983; Phelps, 1972; Aigner and Cain, 1977; Cornell and Welch, 1996; Pinkston, 2003; or Coate and Loury, 1993)

\(^2\) In a recent unpublished paper, Alston (2019) studies individuals’ willingness to pay to hide/include their gender in resumes in an experimental setting. The author finds that, unlike men, women are willing to forfeit part of their salary to hide their gender. This seems rather similar to our results regarding the choice of an avatar.

\(^3\) There are strong differences across gender in the likelihood one pursues STEM careers, with males choosing courses and degrees with a strong mathematical component far more often than do females. These differences emerge at high school (Buser, Niederle and Oosterbeek, 2004; Joensen and Nielse, 2011) and remain at the college level (Zafar, 2013). This gap further widens at the graduate-school level (Hill et al., 2010).
Vesterlund, 2007; Niederle and Vesterlund, 2011) have been extensively analyzed in the literature as an explanation for the self-selection into math-oriented activities. Second, it is possible that men have better aptitudes for science-related tasks; however, the evidence is not conclusive on this matter. Some studies find that males perform better, particularly with respect to being in the top of the distribution than females (Hedges and Nowel, 1995; Xie and Shauman, 2003; Ellison and Swanson, 2010; Guiso, Monte, Sapienza, and Zingales, 2010; Grosse, Riener, and Dertwinkle-Kalt, 2014), while other studies indicating that there are no significant gender differences in performance (Hyde and Mertz, 2009; Hyde et al., 2008; Guiso et al., 2008).

According to Bordalo, Coffman, Gennaioli, and Shleifer (2016), men being over-represented at the highest performance levels leads a stereotypical thinker to exaggerate the magnitude of mean differences. This could potentially lead to the stereotype that male performance in math is high, while female performance is poor.

This paper offers an alternative explanation of why it is less likely for females to choose a STEM major. Rather than focusing on discriminatory behavior, we study experimentally whether men and women react by making strategic decisions when they face a situation of potential discrimination. In particular, we study whether subjects in a labor-market setting hide (or reveal) their gender, perhaps in anticipation of a lower probability of being hired. Moreover, we also study how the decision to hide one’s gender depends on the task. We take this as a starting point of our experimental design, where participants were allocated to a group of seven people and randomly assigned a role as a firm or worker. In each group, there were five workers and two firms. Firms could hire one worker to perform a task in a subsequent stage. The only information that firms had about the workers was an avatar signaling a worker’s gender. Prior to the firms making their decisions, the workers chose the avatar that would represent them in the market.
Each worker could choose: (i) a male avatar, (ii) a female avatar, or (iii) a neutral avatar.\(^4\) Workers’ profits were larger if they were hired by a firm than if they were not, and this was common information to all workers.

In our design, we vary the information that workers receive about the task when they pick the avatar. In the first treatment, workers know that they will be hired to perform a mathematical task that consists of adding five two-digit numbers over a five-minute period. The task is the same in the second treatment, but the workers did not receive information about the task. In the third treatment, workers know that they will be performing a task consisting of identifying emotions depicted on individuals’ faces, which carries a female gender stereotype (Gigerenzer et al., 2013; Halladay, 2017; Bordalo et al., 2016a).

Our study makes two principal contributions. First, we explore whether a potential perception of discrimination can affect the gender of the avatar chosen by people who wish to be hired. Given the stereotype that males are better at math-related tasks and that females are better at emotion recognition, one might expect that male avatars would be the typical choice in the summing-numbers task and that female avatars would be the typical choice in the emotion-recognition task. Second, we demonstrate very clear patterns regarding the performance on these tasks across gender.

Our results show that men and women react differently when they face potential discrimination. When the task is math-oriented and workers have information regarding the task, males do indeed self-identify almost twice as often as females. However, these gender differences disappear in the other two treatments and people largely choose same-gender avatars. When the workers do not know the task prior to the avatar choice, the proportions of male and

---

\(^4\) The use of avatars in an experiment is not new. For example, Fiedler and Haruvy (2009) or Fiedler et al. (2011), conduct experiments in a virtual world to study subjects’ behavior in a Trust Game.
female avatars chosen were identical, at 42.9%. Finally, one might expect males to pretend to be females in the emotion-recognition task, but they do not do so. Although females chose the female avatar in most of the cases (64.1%), the majority of males (67.3%) still picked the male avatar, despite the fact that women are considered to be better at identifying emotions.\(^5\) These results would suggest that most women choose avatars strategically, hiding their gender when the task is male-oriented, but that men do not hide their gender in the female-oriented task. We propose a theoretical model to explain these findings.

Second, we observe interesting results with respect to performance on our tasks. We find a dramatic difference across gender in the adding-numbers task; males (females) average 9.06 (6.00) successful sums. This is indeed consistent with the stereotype and shows an *ex-post* failure by firms to maximize income. In great contrast, we find virtually no difference in mean performance in the emotion-recognition task, with males (females) averaging 12.00 (12.02) successes. This is certainly not consistent with the prevailing stereotype.

Additionally, even considering the limitation of our results regarding the hiring decisions, we see no evidence of discrimination against female avatars. Perhaps surprisingly, firms do not discriminate against female avatars in the math task, despite (female) workers’ expectations. There is also no gender discrimination in the emotion-recognition task. We use our theoretical framework to offer possible explanations that reconcile these results with the findings in terms of avatars’ choices.

The remainder of the paper is organized as follows. In Section 2 we explain the experimental design. Section 3 offers a theoretical framework and presents our hypothesis. Section 4 shows the main results and provides some discussion. We conclude in Section 5.

\(^5\) Our own survey evidence, presented in Appendix G, strongly indicates that the emotions task carries a clear female stereotype in this population.
2. Experimental design and procedures

2.1 Experimental design

The experimental design consists of three treatments: the Math-Information Treatment (MIT, hereafter), the Math-No-Information Treatment (MNIT, hereafter), and the Emotions-Information Treatment (EIT, hereafter). All three treatments involved a real-effort task.

In MIT, participants were allocated into groups of seven people and randomly assigned a role as a firm or a worker. Each group was composed of five workers and two firms. Each firm could hire only one of the five workers, who would then add five two-digit numbers for five minutes (Niederle and Vesterlund, 2007). Participants could not use calculators or scratch paper. Both firms and workers were paid on a piece-rate basis. Firms and workers who were hired were paid 1.5 GBP for each correct sum that workers did. Workers who were not hired by a firm would do the same task, but would only be paid 0.5 GBP per correct sum. We chose adding five two-digit numbers because this task is related to an area for which there is the stereotype that men perform better (Correll, 2001; Rudman et al., 2001; Kiefer and Sekaquaptewa, 2007).

When deciding whom to hire to perform the task, the only information firms had about a worker was an avatar signaling the worker’s gender. Each worker had previously chosen an avatar (male, female, or neutral) that would represent him or her in the market. We chose to use avatars as a key feature of our design because they allow workers to be represented in the labor market by their preferred gender, without having to explicitly state the gender.

---

6 Given the nature of the research question, running an Emotions-No-Information Treatment was not necessary.
7 Workers in the experiment were not explicitly told that the avatar was signaling their gender. They were just told that there would be an avatar that would represent them in the market. See the instructions in Appendix A for more details.
8 See Appendix B for the avatar alternatives.
9 Lim and Harrell (2015) investigated the behavioral patterns of individuals who construct their avatars, including preferences for a particular avatar gender, and looked at the connections between these behavioral patterns and the
When choosing the avatar, workers had information regarding: (i) the task they would perform, (ii) the size and composition of the group (number of firms and workers), (iii) the fact that each firm could only hire one worker, (iv) the payoff scheme and (v) the fact that firms did not know that the avatar was chosen by the worker. Note that workers had not experienced the task when choosing the avatar. The task was performed only once, after the hiring stage was over. Figure 1 provides a summary of all stages in the experiment.

**Figure 1. Structure of the experiment.**

![](image)

MNIT was the same as MIT except that workers had no information about the nature of the task they would perform (in the third stage) when they chose avatars. In this treatment, when choosing avatars, workers had information about: (i) the size and composition of the group (number of firms and workers), (ii) the fact that each firm could only hire one worker, (iii) the participants’ true gender identities. In our design, preferences would be the same across treatments, changing only the strategic considerations of the avatar.

The reason behind offering the possibility of a neutral avatar was to present an intermediate mechanism for a worker to not reveal his or her gender without active misrepresentation. Identity considerations (Akerlof and Kranton, 2000; Charness, Rigotti, and Rustichini, 2007; Chen and Li, 2009) suggest that it is likely to be costly for one to choose an avatar that is not one’s actual gender. We felt that while choosing a neutral avatar was also somewhat costly in relation to one’s identity, this cost would be considerably lower than choosing an avatar from the other gender. This is a standard practice in companies that allow workers to not report their gender, race or religious preferences when applying for a job.

Note that, in this treatment, firms knew the task that workers would perform when hiring them.
fact that their earnings could potentially be three times larger if they were hired than if they were not and (iv) the fact that firms did not know that the avatar was chosen by the worker.

*EIT* was the same as *MIT* with the only difference being that the task performed by the workers in the third stage was not math-related. In *EIT*, workers would perform a real-effort task consisting of identifying emotions from faces. The facial emotion task utilizes professionally classified images obtained from The Great Good Science Center at the University of California, Berkeley.\(^{12}\) Workers were shown 15 faces, each appearing on the screen for two seconds, and subjects attempted to correctly select the depicted emotion out of four options. Subjects had 20 seconds to submit an answer after each image was displayed. They were (accurately) told that the same emotions could repeat but the same image would never appear more than once. As in the other treatments, firms and hired workers would earn 1.5 GBP per correct emotion identified and each non-hired worker would earn 0.5 GBP per emotion. This task was chosen since, contrary to the math task, it carries a gender stereotype that women will perform better than men (Halladay, 2017; Gigerenzer *et al.*, 2013, Bordalo *et al.*, 2016a).

### 2.2 Procedures

The experiment was conducted at the University of Exeter with 434 participants, who were recruited using the online recruitment system ORSEE (Greiner, 2015). The experiment was programmed and conducted with z-Tree (Fischbacher, 2007). We conducted a total of 20 sessions with 21 subjects (6 firms and 15 workers) in each; we also had one session with 14 participants (4 firms and 10 workers). Subjects could see the composition of the session when entering the lab. The gender composition of all sessions was always the same, 50% male

---

\(^{12}\) We thank Brianna Halladay for granting us access to the material she had obtained from the Center.
participants and 50% female participants. No individual participated in more than one session. On average, each person received 10.5 GBP for a session that was 45 minutes or less. Table 1 summarizes the number of sessions and observations for each treatment for each gender.

**Table 1. Experiment summary**

<table>
<thead>
<tr>
<th>Treatment</th>
<th># sessions</th>
<th># workers (male)</th>
<th># workers (female)</th>
<th># workers (total)</th>
<th>#firms (male)</th>
<th>#firms (female)</th>
<th>#firms (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT</td>
<td>7</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>MNIT</td>
<td>7</td>
<td>52</td>
<td>53</td>
<td>105</td>
<td>21</td>
<td>21</td>
<td>42</td>
</tr>
<tr>
<td>EIT</td>
<td>7</td>
<td>52</td>
<td>53</td>
<td>105</td>
<td>21</td>
<td>21</td>
<td>42</td>
</tr>
</tbody>
</table>

To collect information regarding subjects’ beliefs, we conducted a post-experimental questionnaire. Ninety participants, different from those who took part in the experiment, were given the original instructions and were asked to guess the performance of those who took part in the experiment, as well as to identify the level of discrimination faced by male/female workers. They were also asked to provide information regarding their willingness to pay in order to have the chance of being identified as female/male in a labor market setting. Finally, participants in the questionnaire were asked about their preferences towards a particular gender. Subjects’ expected levels of performance, discrimination, willingness to pay, and view regarding represented gender were gathered for each of the three treatments described above. Thirty independent observations were collected for each treatment.

---

13 Note that in each 21-person session there was one extra male or female participant. We chose the gender of the extra participant in each session randomly. However, note that participants were told that they would be allocated to groups of seven subjects. So, even if they could see the gender composition of the session, it was impossible for them to know the gender composition of their group.

14 See Appendix A11 for the complete questionnaire.
3. Model and Hypotheses

In this section, we offer a simple model that captures the main features of our experimental design and that allow us to state our main hypotheses.\(^{15}\)

Consider an economy populated by three workers – \(A_i\), for \(i=1,2,3\) – and one employer \(P\), who chooses to hire one worker to perform a task \(t \in \{t_x, t_y\}\). Each worker is characterized by a gender \(g_i \in \{x, y\}\). In what follows, we set \(g_1 = g_3 = x\), and \(g_2 = y\). The task \(t_x\) (respectively, \(t_y\)) carries a gender-\(x\) (respectively, gender-\(y\)) stereotype, that is, \(P\) believes that workers whose gender is \(x\) (respectively, \(y\)) perform better than workers whose gender is \(y\) (respectively, \(x\)) at solving this task.\(^{16}\) We assume that each worker’s gender is common knowledge.\(^{17}\) This simplified setting captures the main features of our experimental design, in which two employers can hire one worker (each) from a pool of five candidates.

Each worker \(A_i\) reports his or her gender to \(P\), where \(A_i\)’s report \(r_i \in \{x, y\}\) can either be truthful \((r_i = g_i)\) or not \((r_i \neq g_i)\), for \(i = 1,2,3\).\(^{18}\) \(A_i\) bears a cost \(C_{g_i} \geq 0\) from misreporting her/his true gender. After observing \(\{r_1, r_2, r_3\}\), \(P\) chooses whether to hire \(A_i\) – that is, \(a_i \in \{0,1\}\), where \(a_i = 1\) (respectively, \(a_i = 0\)) if \(A_i\) is hired (respectively, not hired). Formally, \(P\) sets

\(^{15}\) We thank the editor and an anonymous reviewer for suggesting that we develop a formal model that generates the hypotheses that we present later in this section.

\(^{16}\) In the post-experimental survey, we collected participants’ beliefs regarding male/female performance for both tasks (see Table 8 in Appendix J for an overview of the survey results). Results support our assumptions. In the mathematical task, the expected performance was 15.57 for males and 15.07 for females. Hence, participants expect male students to perform better than female students in the numbers task, although differences are only weakly significant \((t = 1.595, p = 0.061,\) one-tailed paired \(t\)-test). In the emotions task, the expected performance is 9.93 for males and 11.60 for females. Differences are statistically significant \((t = 5.281, p < 0.001,\) one-tailed paired \(t\)-test).

\(^{17}\) Note that, in our experimental design participants know the gender distribution in the population of participants (i.e., they observe the gender distribution in a session) but do not know the gender distribution in their own group. Assuming that each player observes the distribution of workers’ gender in her/his own group simplifies our analysis without affecting our main results.

\(^{18}\) In line with our experimental design, Appendix I considers an extension of the model in which workers can choose between truthfully reporting their gender, misreporting their gender, or not reporting their gender (i.e., reporting a “neutral” gender). Allowing workers to report a “neutral” gender does not affect the main results reported in this section.
\( \{a_1, a_2, a_3\} \), where \( \sum_{i=1}^{3} a_i = 1 \). If hired, \( A_i \) gains \( \bar{u} \), whereas she/he gains \( u \) if not hired, with \( \Delta u = \bar{u} - u > 0 \). Given \( g_i \), \( A_i \) ’s (ex-ante) utility is:

\[
U_i(g_i) = a_i\bar{u} + (1 - a_i)u - C_{g_i}1_{r_i \neq g_i},
\]

for \( i = 1,2,3 \), where \( 1_{r_i \neq g_i} \) is an indicator function that takes value 1 if \( A_i \) reports a gender that differs from her/his actual gender.

To capture one of the main features of our experimental design – namely P’s unawareness about \( A_i \) ’s ability to manipulate her/his reported gender (i.e., the avatar) – we assign P to one of three different types. Formally, P has a type \( \theta \in \{\theta_1, \theta_2, \theta_3\} \), which is unobservable to the workers. Type \( \theta_1 \) hires a worker who reports \( r_i = y \) (if any do). If more than one worker reports \( r_i = y \), type \( \theta_1 \) hires each of these workers with equal probability. Similarly, if \( r_i = x \), \( \forall i \), type \( \theta_1 \) hires each worker with a probability equal to \( 1/3 \). Type \( \theta_2 \) hires a worker who reports \( r_i = x \) (if any do). If more than one worker reports \( r_i = x \), type \( \theta_2 \) hires each of these workers with equal probability. Similarly, if \( r_i = y \), \( \forall i \), type \( \theta_2 \) hires each of these workers with a probability of \( 1/3 \). Finally, type \( \theta_3 \) has no preference over workers’ gender. Type \( \theta_3 \) hires each worker with a probability equal to \( 1/3 \), independently of the workers’ reported gender.

We denote by \( \beta_y, \beta_x \) and \( 1 - \beta_y - \beta_x \), the probabilities that \( \theta = \theta_1, \theta = \theta_2 \) and \( \theta = \theta_3 \), respectively. The probability distribution of \( \theta \) depends on the task \( t \in \{t_x, t_y\} \) to be performed, which is known to P. In particular, if \( t = t_y \), based on the gender-\( y \) stereotype carried by this task, we assume \( \beta_y > \beta_x = 0 \). Likewise, based on the gender-\( x \) stereotype carried by this task, if \( t = t_x \), we assume \( \beta_x > \beta_y = 0 \). In the following, we assume \( t = t_y \) in MNIT.

---

\(^{19}\) We abstract from the workers’ effort task. Given our modeling of P’s behavior (see below), this choice has no consequences on our main hypotheses and findings.
We denote by $\beta_y$, $\beta_x$ and $1 - \beta_y - \beta_x$, workers’ expected probabilities that $\theta = \theta_1$, $\theta = \theta_2$ and $\theta = \theta_3$, respectively. Unlike $P$, workers may ($MIT$ and $EIT$) or may not ($MNIT$) have information about the task to be performed. When workers are informed about $t = t_y$, we assume $\hat{\beta}_y > \hat{\beta}_x = 0$. Likewise, when workers are informed about $t = t_x$, we assume $\hat{\beta}_x > \hat{\beta}_y = 0$. In other words, when informed about the task to be performed, workers believe that firms carry a gender stereotype depending on the task to be performed. The subjects’ beliefs we collected in the post-experimental questionnaire (see Table 8 in Appendix J) confirm that $\beta_y > \beta_x$ (respectively, $\beta_x > \beta_y$) when workers are informed about $t = t_y$ (respectively, $t = t_x$). Finally, when workers are uninformed about the task to be performed (and about the set of available tasks), we set $\beta_y > \beta_x = 0$. Assuming $\beta_y > \beta_x$ is in line with our post-experimental questionnaire, which suggests that participants anticipate discrimination when the task is unknown.\(^{20,21}\)

Players interact for one period. In $MNIT$, the timing of the game is:

1. Given $t = t_y$, nature draws $\theta$;
2. $A_i$ does not observe $t$, for $i = 1, 2, 3$. Workers simultaneously choose $r_i \in \{x, y\}$, for $i = 1, 2, 3$;
3. $P$ observes $\{r_1, r_2, r_3\}$ and, given $t = t_y$, chooses $\{a_1, a_2, a_3\}$;
4. Payoffs are realized.

\(^{20}\) In the post-experimental survey, we find that subjects believe that firms would hire a majority of male workers (56.87%) in the mathematical task (see Table 8, Panel A, in Appendix J). The opposite (43.13% of male workers hired) is true in the emotions task: Beliefs are significantly different from 50% both in the math ($t = 3.221, p = 0.003$, two-tailed $t$-test) as well as the emotions ($t = -2.704, p = 0.011$, two-tailed $t$-test) task. Finally, when subjects do not know the task that will be performed, they anticipate that a majority of male workers (55.29%) will be hired. This percentage is significantly different from 50% ($t = 3.872, p < 0.001$, two-tailed $t$-test). As shown in Panel B and Panel C in Table 8 (in Appendix J), these findings remain virtually unchanged when we distinguish between female and male respondents to our survey.

\(^{21}\) Although assuming $t = t_y$ and $\hat{\beta}_y > \hat{\beta}_x = 0$ in $MNIT$ may induce the reader to attach a masculine gender to $g = y$, genders $x$ and $y$ can be interchangeably interpreted as the masculine or the feminine gender in our theoretical analysis.
In MIT and EIT, the timing of the game is:

1. Given \( t \in \{ t_x, t_y \} \), nature draws \( \theta \);
2. \( A_i \) observes \( t \), for \( i = 1, 2, 3 \). Workers simultaneously choose \( r_i \in \{ x, y \} \), for \( i = 1, 2, 3 \);
3. \( P \) observes \( \{ r_1, r_2, r_3 \} \) and, given \( t \in \{ t_x, t_y \} \), chooses \( \{ a_1, a_2, a_3 \} \);
4. Payoffs are realized.

We denote \( A_1 \)'s probability of playing \( r_1 = x \) by \( p \). \( A_2 \)'s probability of playing \( r_2 = x \) is denoted by \( q \), and \( A_3 \)'s probability of playing \( r_3 = x \) is denoted by \( s \). Given \( g_1 = x \), from (1), \( A_1 \)'s (ex-ante) utility \( U_i(g_i, r_i) \) from truthfully reporting \( r_1 = x \) is:

\[
U_1(x, x) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right]
+ \beta_y \left[ (s(1 - q) + q(1 - s)) \frac{1}{2} \bar{u} + \frac{1}{2} u \right]
+ \beta_x \left[ (s(1 - q) + q(1 - s)) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + s q \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right]
+ (1 - q)(1 - s) \bar{u},
\]

(2)

whereas \( A_1 \)'s (ex-ante) utility from (mis)reporting \( g_1 \), i.e., \( r_1 = y \), is:

\[
U_1(x, y) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right]
+ \beta_y \left[ (s(1 - q) + q(1 - s)) \frac{1}{2} \bar{u} + \frac{1}{2} u \right]
+ \beta_x \left[ (s(1 - q) + q(1 - s)) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + s q \bar{u} \right]
+ \beta_y \left[ (s(1 - q) + q(1 - s) + s q) \bar{u} + (1 - q)(1 - s) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] - C_x.
\]

(3)

\( A_2 \) and \( A_3 \)'s utilities – conditional on \( (g_i, r_i) \), for \( i = 2, 3 \) – are similarly defined. We also define the (ex-ante) probability that \( P \) hires a worker whose reported gender is \( x \):

\[
H_x = p q s + [q s (1 - p) + p q (1 - s) + p s (1 - q)] \left[ \beta_x + \frac{2}{3} \left( 1 - \beta_y - \beta_x \right) \right] +
\]

12
\[ s(1-p)(1-q) + q(1-p)(1-s) + p(1-q)(1-s) \left[ \beta_x + \frac{1}{3} (1 - \beta_y - \beta_x) \right]. \quad (4) \]

In line with the treatments considered in our experiment, we proceed by analyzing the Nash equilibria of the game in the following two cases: (i) \( \hat{\beta}_y > \hat{\beta}_x = 0 \) and (ii) \( \hat{\beta}_x > \hat{\beta}_y = 0 \). All proofs are relegated to Appendix H.

### 3.1 MNIT

Consider first the game in which the task to be performed \( t = t_y \) is known to \( P \) but not to workers. Recall that, in accordance with the findings in our post-experimental survey (Table 8 in Appendix J), we set \( \hat{\beta}_y > \hat{\beta}_x = 0 \) in MNIT.

**Proposition 1.** Suppose \( \hat{\beta}_y > \hat{\beta}_x = 0 \). Then, at equilibrium \( A_2 \) truthfully reports her/his gender, i.e., \( r_2 = y \). Also:

1. If \( \frac{c_x}{\beta_y} < \frac{1}{3} \) \( A_1 \) and \( A_3 \) misreport their true gender, i.e., \( r_1 = r_3 = y \);
2. If \( \frac{c_x}{\beta_y} \in \left[ \frac{1}{3}, \frac{1}{2} \right) \), three equilibria exist in which either \( \{r_1, r_3\} = \{x, y\} \), or \( \{r_1, r_3\} = \{y, x\} \), or both \( A_1 \) and \( A_3 \) report \( x \) with probabilities \( p^* = s^* = 2 \left[ \frac{3c_x}{\beta_y} - 1 \right] \);
3. If \( \frac{c_x}{\beta_y} \geq \frac{1}{2} \) \( A_1 \) and \( A_3 \) truthfully report their gender, i.e., \( r_1 = r_3 = x \).

In MNIT, subjects’ anticipate discrimination despite the fact that the task to be performed is unknown to them. In our simple framework, Proposition 1 suggests that workers whose gender can potentially be negatively affected by discrimination react by hiding their gender, provided that (i) this is not too costly and (ii) the anticipated level of discrimination is sufficiently strong (i.e., \( \beta_y \) is sufficiently high).

---

22 Given \( g_1 = g_3 = x \), and \( g_2 = y \), we consider both cases (i) and (ii) to show that our main results in MIT and EIT are not affected by workers’ gender composition.

23 All the equilibria presented in Sections 3.1 and 3.2 hold in a more general setting in which \( \hat{\beta}_g > \hat{\beta}_{-g} > 0 \), with \( \hat{\beta}_g = \hat{\beta}_{-g} + \delta \) and \( \beta_g + \beta_{-g} \leq 1 \), for \( \delta \) positive and sufficiently large, and where \( g, -g \in \{x, y\} \), and \( g \neq -g \).
3.2 MIT and EIT

Consider now the cases in which all players know the task.

**Proposition 2.** Suppose \( t = t_y \). Given \( \hat{\beta}_y > \hat{\beta}_x = 0 \) and \( \beta_y > \beta_x = 0 \), the equilibrium behavior described in Proposition 1 holds.

**Proposition 3.** Suppose \( t = t_x \). Given \( \hat{\beta}_x > \hat{\beta}_y = 0 \) and \( \beta_x > \beta_y = 0 \), at equilibrium \( A_1 \) and \( A_3 \) truthfully report their gender, i.e., \( r_1 = r_3 = x \). Also:

- a. if \( \frac{c_y}{\bar{\beta}_x \Delta u} < \frac{1}{3} \), \( A_2 \) misreports her\'his true gender, i.e., \( r_2 = x \);
- b. if \( \frac{c_y}{\bar{\beta}_x \Delta u} \geq \frac{1}{3} \), \( A_2 \) truthfully reports her\'his gender, i.e., \( r_2 = y \).

As with Proposition 1, Propositions 2 and 3 suggest that workers who face potential discrimination react by hiding their gender, provided that (i) this is not too costly and (ii) the anticipated level of discrimination is sufficiently strong. If these two conditions hold, Propositions 1, 2 and 3 allow us to state the following hypotheses:

**Hypothesis 1:** In MNIT, male workers are more likely to truthfully report their gender than are female workers.

**Hypothesis 2:** In MIT, male workers are more likely to truthfully report their gender than are female workers.

**Hypothesis 3:** In EIT, female workers are more likely to truthfully report their gender than male workers.

Hypothesis 1 reflects the general belief that there is discrimination against female workers, so that it would be strategic to be seen as male. It turns out that this belief is also reflected in the data from our survey of the student population. Hypotheses 2 and 3 reflect the idea that workers will strategically wish to be seen as the gender that is considered to be better at the respective task.
4. Results

This section is structured as follows. We first analyze the avatar choice. Second, we look at firms’ decisions regarding whom to hire. We then study workers’ performance and consider worker profits in relation to avatar choices. Finally, we provide some discussion of our findings.

Table 2 summarizes the avatar choices made by the workers. Specifically, we report the percentage of workers choosing a male, a female, or a neutral avatar. We perform this analysis for both male and female workers.

Table 2. Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>MIT Avatar</th>
<th></th>
<th>MNIT Avatar</th>
<th></th>
<th>EIT Avatar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Neutral</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Males</td>
<td>68.00%</td>
<td>10.00%</td>
<td>22.00%</td>
<td>59.62%</td>
<td>21.15%</td>
</tr>
<tr>
<td>Females</td>
<td>40.00%</td>
<td>36.00%</td>
<td>24.00%</td>
<td>26.42%</td>
<td>64.15%</td>
</tr>
</tbody>
</table>

How to read this Table: The leftmost column shows that, in the MIT treatment, 68% of all males chose the Male avatar, while 40% of all females chose the Male avatar. The middle column shows that, in the MNIT treatment, 21.15% of all males chose the Female avatar, while 64.15% of all females chose the Female avatar. The proportion for female avatar choice by females in MNIT and EIT is in fact identical.

4.1. Avatar choices

In this section we consider the avatar choices made by each gender.

4.1.1 Numbers task

In this subsection, we consider workers’ avatar choices in the adding-numbers task according to whether the task is known or unknown.

The middle panel in Table 2 shows that there is no significant difference in self-identification rates across gender ($Z = -0.476$, $p = 0.634$, Mann Whitney test) when workers do
not have any information regarding the math-related task, with 59% of males (64% of females) choosing a male (female) avatar. This result contrasts with our Hypothesis 1. In this treatment, we also observe some differences in the choice of neutral avatars. Males chose a neutral avatar twice as often as females, 19% versus 9%, but this difference is not statistically significant ($Z = 1.427, p = 0.153$, Mann-Whitney test).

**Result 1:** When workers do not know the task, there are no significant differences between the avatar-choosing behavior of male and female workers. Neither the self-identification rates nor the percentage of neutral avatars chosen are significantly different across genders.

In MIT, workers know that the task to be performed is math-related and matters change dramatically. As shown in the left panel in Table 2, although the percentage of males choosing a male avatar goes up to 68%, this increment is not statistically significant ($Z = 0.876, p = 0.381$, Mann-Whitney test). However, for female workers we observe a deep reduction in the self-identification rate, which drops from 64% in MNIT to 36% in MIT. This difference is strongly significant ($Z = -2.842, p = 0.004$, Mann-Whitney test). As a result, a significantly larger percentage of male workers report their true gender in MIT ($Z = -3.187, p = 0.001$, Mann-Whitney test). These findings support our Hypothesis 2.

Moreover, in MIT, female workers choose a male avatar 40% of the time, which is a 50% increase with respect to MNIT; they also choose the neutral avatar 24% of the times, which is twice as frequent as in MNIT. While the difference between treatments is not statistically significant for male avatar choices ($Z = 1.458, p = 0.145$, Mann-Whitney test), differences are significant for the choice of neutral avatars ($Z = 1.981, p = 0.048$, Mann-Whitney test). These

---

24 Here and elsewhere, we round $p$-values to the third decimal place; tests are two-tailed unless otherwise indicated. 25 The proportion of female (10%) and neutral (22%) avatars chosen by males in MIT is also not significantly different from that in MNIT ($Z = -1.541, p = 0.123$, and $Z = 0.344, p = 0.731$, Mann-Whitney test for the choice of female and neutral avatars, respectively).
results show that when female workers decide not to report their true gender, they do not always choose the male avatar, but sometimes instead opt for the alternative of hiding their gender without explicitly making a false indication.

**Result 2:** *When it is common information that the task is math-oriented, male workers do not change their behavior significantly compared to the control treatment. Female workers, however, significantly decrease (increase) the proportion of female (neutral) avatars chosen.*

4.1.2 Emotion-recognition task

In this subsection, we analyze workers’ avatar choices when the task is more female-oriented. As shown in Table 2, the choice of the avatar in *EIT* is similar to that in *MNIT*, when workers do not have information about the task. In *EIT*, 67.30%, 13.46% and 19.23% of males choose the male, female and neutral avatars, respectively; these figures are 59.62%, 21.15%, and 19.23% in *MNIT*. None of the pairwise differences across the two treatments is statistically significant (*Z* = -0.811, 1.032, and 0.000, *p* = 0.418, *p* = 0.302, and *p* = 1.000, respectively, Mann-Whitney test for the proportion of male, female and neutral avatars chosen).

A similar result is found for female workers’ choices. In *MNIT*, the percentage of women choosing a male, female and neutral avatar is 26.42%, 64.15% and 9.43%, respectively. The equivalent percentages in *EIT* are identical, so that here there are no significant differences across treatments either (*Z* = 0.000, *p* = 1.000, Mann-Whitney test). As a result, a similar percentage of male and female workers report their true gender in *EIT* (*Z* = -0.339, *p* = 0.734, Mann-Whitney test). This contradicts our Hypothesis 3.

**Result 3:** *When the task is female oriented both male and female workers behave similarly to the case with no task information.*

Next, we provide an econometric analysis to better understand workers’ avatar decisions in the different tasks. Table 3 reports the results of a multinomial logit model, to study the
likelihood of subjects’ choosing a true, neutral or opposite avatar (opposite avatar served as a reference category). We use the following explanatory variables: *Male*, a binary covariate that equals 1 if the worker is male, and 0 otherwise; *STEM*, a dummy with the value 1 if the worker is currently studying a degree in Science, Technology, Engineering or Mathematics, and 0 otherwise; and *Treatment1 (Treatment3)*, with value 1 if *MIT (EIT)* and 0 otherwise. We also add an interaction between *Male* and *Treatment1* in order to determine whether the information about the task that workers receive prior to choosing the avatar affects males and females differently.

**Table 3. Multinomial logit regressions on reporting true, neutral and opposite gender**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>truthful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>0.974***</td>
<td>(0.325)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>0.142</td>
<td>(0.475)</td>
<td></td>
</tr>
<tr>
<td>STEM</td>
<td></td>
<td>-0.475</td>
<td>(0.344)</td>
<td></td>
</tr>
<tr>
<td>Treatment1</td>
<td></td>
<td>-0.999**</td>
<td>(0.461)</td>
<td></td>
</tr>
<tr>
<td>Treatment3</td>
<td></td>
<td>0.000</td>
<td>(0.451)</td>
<td></td>
</tr>
<tr>
<td>Male*Treatment1</td>
<td></td>
<td>1.939***</td>
<td>(0.752)</td>
<td></td>
</tr>
<tr>
<td>Male*Treatment3</td>
<td></td>
<td>0.606</td>
<td>(0.707)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>-0.951*</td>
<td>(0.527)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>0.928</td>
<td>(0.681)</td>
<td></td>
</tr>
<tr>
<td>STEM</td>
<td></td>
<td>-0.418</td>
<td>(0.443)</td>
<td></td>
</tr>
<tr>
<td>Treatment1</td>
<td></td>
<td>0.564</td>
<td>(0.639)</td>
<td></td>
</tr>
<tr>
<td>Treatment3</td>
<td></td>
<td>0.000</td>
<td>(0.737)</td>
<td></td>
</tr>
<tr>
<td>Male*Treatment1</td>
<td></td>
<td>0.373</td>
<td>(0.943)</td>
<td></td>
</tr>
<tr>
<td>Male*Treatment3</td>
<td></td>
<td>0.481</td>
<td>(0.989)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td>309</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>-278.179</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. ***, **, and * denote significance at \( p = 0.01, 0.05, \) and 0.10, respectively. The opposite avatar served as the reference category.
First, we examine the probability of reporting the real avatar with respect to choosing the opposite avatar. The estimate of Male shows that there are no general gender differences in the probability of choosing the avatar that reveals the true gender of the worker. Also, we observe that being in the treatment in which workers know that they will be doing the math task decreases the probability of reporting the real avatar, compared to the case in which workers do not receive any information regarding the task. However, this is not true for the emotions-recognition task. Subjects do not reveal their true gender significantly less in EIT than in MNIT. Finally, results show that releasing information about the task affects male and female workers differently, depending on the task. While males report their true gender with a higher probability when the task is math-oriented, we observe no gender effect in revealing the worker’s true gender in the emotion-recognition task.

We then focus on the likelihood of choosing a neutral avatar. For this case, Table 3 shows that neither the gender of the worker nor the type of task that workers will perform affect the probability of subjects reporting a neutral avatar over the opposite one.

4.2. Hiring decisions

We now discuss the behavior of the firms. Note that, as a consequence of the endogenous choice of avatars, the distribution of candidates that firms face can differ across sessions. It could happen that particular distributions affect firms’ decisions when it comes to hire one worker. Given the limitation of our sample for the different distributions, our results should be seen more as initial evidence than as very reliable findings.

As a first cut we observe that, in MIT, the percentage of male and female avatars hired (as a proportion of all hired avatars) is 37.50 and 35.00, respectively. The equivalent figures are
38.09 and 40.47 in MNIT, and 38.09 and 50.00 in EIT.\textsuperscript{26} This result indicates that firms hire a similar proportion of male and female avatars regardless of the treatment.\textsuperscript{27}

Moreover, the proportion of male and female avatars hired by male and female firms is also not dramatically different. Pooling the numbers data from MIT and MNIT, we find that the percentage of male avatars and female avatars hired by male firms is 37.50\% and 30.00\%, respectively. Differences are not statistically significant (\(p = 0.478\), test of proportions). The respective figures for the percentage of male and female avatars hired by female firms are 38.09\% and 45.24\%; this difference is not statistically significant (\(p = 0.507\), test of proportions).

There are also no dramatic differences in hiring decisions in the emotion-recognition task (EIT). The percentages of male and female avatars hired are 38.09\% and 42.86\% when the firm is male and 38.09\% and 57.14\% when the firm is female. Differences are not statistically significant for either case (\(p = 0.753\) and \(p = 0.216\), test of proportions for male and female firms, respectively). Overall, this analysis indicates that there is no discrimination against female avatars in any treatment.\textsuperscript{28}

**Result 4:** There is no discrimination against female avatars in either treatment.

To better understand firms’ behavior, we next present individual level analysis of hiring decisions. The first column in Table 4 reports the marginal effects from a Probit model, where the dependent variable is Male Avatar Hired, a dummy taking the value 1 if the firm hires a male avatar and 0 otherwise. We use the following explanatory variables: Male Firm, a binary covariate that equals 1 if the firm is male, and 0 otherwise; Treatment1 (Treatment3), taking

\textsuperscript{26} Differences are not statistically significant in any treatment (\(Z = 0.233, p = 0.816\), \(Z = -0.223, p = 0.823\), and \(Z = -1.099, p = 0.271\) for MIT, MNIT, and EIT, respectively, all tests of proportions). The power of the tests is 0.818, 0.825, and 0.401 for MIT, MNIT, and EIT, respectively.

\textsuperscript{27} As we do not discuss the hiring results with neutral avatars in the text, we provide these in Appendix E.

\textsuperscript{28} Alternative analyses on hiring decisions are reported in Appendix D and lead to the same conclusions.
value 1 if MIT (EIT) and 0 otherwise; and Number of males (neutrals), is the number of male (neutral) avatars observed by the firm. We also add an interaction between Male Firm and Treatment1 and an interaction between Male Firm and Treatment3.

Column (2) repeats the specification in column (1) but in this case using Female Avatar Hired as the dependent variable; this is a dummy taking the value 1 if the firm hires a female avatar and 0 otherwise.

Table 4 shows that the gender of the firm does not affect the probability of hiring a male or a female avatar. We also observe that the task to be performed in the second stage of the experiment does not have an effect on the probability of hiring a particular avatar. Results show that the number of male avatars available in the group increases (decreases) the probability that the firm hires a male (female) avatar. Finally, the number of neutral avatars that are present only affects the probability that a female avatar is hired.

Table 4. Probit regression on the probability of hiring a male (female) avatar

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Firm</td>
<td>-0.085</td>
<td>-0.168</td>
</tr>
<tr>
<td></td>
<td>(0.150)</td>
<td>(0.157)</td>
</tr>
<tr>
<td>Treatment1</td>
<td>-0.210</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>(0.146)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>Treatment3</td>
<td>-0.088</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>(0.149)</td>
<td>(0.161)</td>
</tr>
<tr>
<td>Number of males</td>
<td>0.167***</td>
<td>-0.200***</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>Number of neutrals</td>
<td>0.031</td>
<td>-0.295***</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>Male Firm*Treatment1</td>
<td>0.201</td>
<td>-0.062</td>
</tr>
<tr>
<td></td>
<td>(0.227)</td>
<td>(0.232)</td>
</tr>
<tr>
<td>Male Firm*Treatment3</td>
<td>0.090</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.223)</td>
<td>(0.225)</td>
</tr>
<tr>
<td>Observations</td>
<td>124</td>
<td>124</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-76.929</td>
<td>-69.788</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. ***, **, and * denote significance at \( p = 0.01, 0.05, \) and 0.10, respectively.
4.3. Workers’ performance

This section analyzes the main determinants of workers’ performance in the real-effort task. Table 5 shows the mean performance for males and females in each of our treatments.

Casual inspection shows that the mean performance is quite different in the numbers task but practically identical with the emotions task. We can also explore the distribution of performance, as shown in Figures 4 and 5. In Figure 4, the Kolmogorov-Smirnov test of cumulative distributions finds a very strong difference in the addition task, $X^2_2 = 20.589$, $p = 0.000$. We observe clear first-order stochastic dominance. Consistent with some previous results, we do see a significant difference ($Z = 3.152$, $p = 0.002$) at the high end of the distribution – 15 of 102 males (14.7%) scored 14 or higher compared to 3 of 103 females (2.9%).

Table 5. Performance by task and gender

<table>
<thead>
<tr>
<th></th>
<th>Numbers</th>
<th>Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>9.059</td>
<td>12.000</td>
</tr>
<tr>
<td></td>
<td>(0.453)</td>
<td>(0.296)</td>
</tr>
<tr>
<td>Female</td>
<td>6.000</td>
<td>12.019</td>
</tr>
<tr>
<td></td>
<td>(0.347)</td>
<td>(0.223)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses.

Figure 4: Cumulative distribution of performance in the math task, by gender
In Figure 5, the Kolmogorov-Smirnov test finds no serious difference in the emotions task, $X^2_2 = 2.580, p = 0.275$. Recall that the mean of the number correct was essentially identical (12) for males and females on the emotion-recognition task. While the differences in Figure 5 are small, they suggest that there is second-order stochastic dominance for females over males, which is consistent with the smaller variance for males shown in Table 4. Thirty-five women (34.0%) scored within one point of the mean (11, 12, or 13) compared to 23 of the men (22.5%).

**Figure 5. Cumulative distribution of performance in the emotions task, by gender.**

![Graph showing cumulative distribution of performance in emotions task by gender.](image)

We now proceed with econometric analysis. Column (1) in Table 6 uses the data from MIT and MNIT and reports the coefficients of an OLS model in which the dependent variable is the number of correct sums solved by the worker. We include as predictors of the performance the same set of explanatory variables used in the analysis of the probability of self-identification (Male, STEM and Treatment). We add to the model a dummy for Being Hired, taking the value 1 if the worker was hired, and 0 otherwise; and Real Avatar, a dummy with value 1 if one reported one’s true gender in the first stage of the game, and 0 otherwise. We also add the interaction between Real Avatar and Being Hired, allowing us to assess whether subjects who were hired after hiding their gender behaved differently from those who truthfully reported it.
Individual performance is only affected by the gender of the worker, with males solving significantly more sums than females. This gender difference could be related to the fact that, in line with previous literature (Xie and Shauman, 2003; Ellison and Swanson, 2010; Bordalo et al., 2016), 83% of the workers who are performing in the top 10% of the distribution were male. Having a larger piece rate when the worker is hired does not affect one’s performance. The interaction between Being Hired and Real Avatar is also not significant, so it seems that workers who were hired after not self-identifying do not perform any differently in the task. This suggests that even a small piece rate was a sufficiently strong incentive to induce workers to exert maximum effort, leaving no room for improvement when incentives were higher.

Table 6. OLS regression on performance

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.358***</td>
<td>4.295***</td>
<td>11.917***</td>
</tr>
<tr>
<td></td>
<td>(0.714)</td>
<td>(0.648)</td>
<td>(0.478)</td>
</tr>
<tr>
<td>Male</td>
<td>3.132***</td>
<td>2.897***</td>
<td>-0.077</td>
</tr>
<tr>
<td></td>
<td>(0.583)</td>
<td>(0.600)</td>
<td>(0.387)</td>
</tr>
<tr>
<td>STEM</td>
<td>0.999</td>
<td>0.862</td>
<td>-0.083</td>
</tr>
<tr>
<td></td>
<td>(0.705)</td>
<td>(0.707)</td>
<td>(0.483)</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.857</td>
<td>0.960*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.580)</td>
<td>(0.579)</td>
<td></td>
</tr>
<tr>
<td>Being Hired</td>
<td>-0.282</td>
<td>0.477</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>(0.880)</td>
<td>(0.591)</td>
<td>(0.657)</td>
</tr>
<tr>
<td>Real Avatar</td>
<td>-0.322</td>
<td>0.339</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.767)</td>
<td>(0.513)</td>
<td></td>
</tr>
<tr>
<td>Being Hired*Real Avatar</td>
<td>1.201 (1.188)</td>
<td></td>
<td>-0.378 (0.811)</td>
</tr>
<tr>
<td>Male Avatar</td>
<td></td>
<td>0.613</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.617)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>204</td>
<td>204</td>
<td>105</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.119</td>
<td>0.123</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Notes: Standard errors are in parentheses. ***, **, and * denote significance at $p = 0.01, 0.05, \ and 0.10$, respectively.

Column (2) replaces Real Avatar with Male Avatar, a binary covariate that takes value 1 if the avatar chosen by the worker was male, and 0 otherwise. Results for this second
specification are the same as in Column (1), with the only significant effect coming from the Male dummy. Column (3) repeats the specification in column (1) using the data from EIT instead. In this case, the dependent variable is the number of emotions correctly identified. Results are very similar to those in the mathematical task with the only difference being that in this case the gender of the workers does not explain their performance.

**Result 5:** Only the gender of the worker affects his/her performance in the math task, with males being more productive than females. The gender does not affect workers’ performance in the emotion-recognition task.

Our data strongly indicate that there is a gender difference in performance in adding up two-digit numbers. While this conflicts with the prevailing view, with a significance level of $p < 0.001$ our results hardly seem accidental. One potential explanation for our findings is what the psychological literature terms stereotype threat, which basically means that the activation of a specific stereotype may negatively impact the task performance of the negatively-stereotyped group (Steele & Aronson, 1995; Ryan & Ryan, 2005; Günther et al., 2010).

The subjects in the experiment were not unusual in any way and people have not historically been suspicious of results found in a standard subject pool of British students (in this case Exeter). So we stand by our result, which is certainly consistent with the near-universal stereotype. It seems important for researchers to report (and publish) non-conforming outcomes reached through standard experimental methodology.

Perhaps surprisingly, the nearly-identical performance in the emotion-recognition task is not consistent with the stereotype that women are better at this task. So hiring female avatars is at least not costly (in expectation) here for the hiring firm. Halladay (2017) finds a similar result with the identical task: Males correctly identify an average of 8.19 emotions, while females correctly identify 8.08 emotions ($p = 0.806$). So support for this stereotype seems elusive.
4.4. Avatar choices and worker’s earnings

What earnings did workers receive in each treatment when choosing different avatars? Table 7 shows workers’ earnings by gender and avatar.\textsuperscript{29} Males did best by choosing female avatars in the MIT and EIT treatments and by choosing male avatars in the MNIT treatment. The difference approaches significance only with MIT, however (12.00 versus 6.43, \( p = 0.073 \)).\textsuperscript{30,31}

<table>
<thead>
<tr>
<th>Table 7. Worker’s earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male workers</strong></td>
</tr>
<tr>
<td>Male avatar</td>
</tr>
<tr>
<td>Neutral avatar</td>
</tr>
<tr>
<td>Female avatar</td>
</tr>
</tbody>
</table>

| **Female workers**           | **MIT** | **MNIT** | **EIT** |
| Male avatar                 | 4.82 \([20]\) (0.96) | 5.11 \([14]\) (0.95) | 12.03 \([14]\) (1.63) |
| Neutral avatar              | 4.71 \([12]\) (0.73) | 11.20 \([5]\) (4.35) | 10.00 \([5]\) (2.98) |
| Female avatar               | 5.97 \([18]\) (1.20) | 5.20 \([34]\) (0.72) | 12.35 \([34]\) (1.07) |

Note: Standard errors are in parentheses. The number of observations is in italics in brackets.

Female workers also did best by choosing female avatars in the MIT and EIT treatments, but a neutral avatar was the optimal choice in the MNIT treatment. In no case was it qualitatively advisable for a female worker to choose a male avatar over a female avatar.

\textsuperscript{29} Workers earn more with the emotion-recognition task than the adding-sums task, but this simply reflects the payoff calibration in the different treatments.

\textsuperscript{30} We have \( p = 0.474 \), and \( p = 0.508 \) for the male comparisons in MNIT, and EIT, respectively.

\textsuperscript{31} Note that in some cells there are only a few observations, so that the corresponding statistical tests are weak.
4.5. Discussion

In this section, we discuss several potential explanations for the discrepancies between experimental results and theoretical predictions regarding workers’ avatar choices. We also briefly discuss the firms’ hiring behavior.

Our results show that female participants behaved as predicted by the theory in two treatments (MIT and EIT). Similarly, male participants also behaved as predicted in two out of three treatments (MNIT and MIT). However, in terms of subjects’ avatar choices, in contrast to our predictions, the majority of female workers chose a female avatar in MNIT. Also, in contrast to what was predicted, the majority of male workers chose a male avatar in EIT. In other words, although female respondents in our post-experimental survey anticipated potential discrimination both when the task was unknown and when facing a task carrying a male stereotype (see Panel C in Table 8, Appendix J), female workers strategically changed their behavior only when informed about the task. By contrast, male workers behaved no differently independently of the task performed and independently of whether the task was known to them.

Finally, in terms of hiring behavior, because both tasks carry a gender stereotype, our theoretical analysis predicts that the proportion of hired male and female avatars differs in all treatments. However, we found no evidence of discrimination at the hiring stage (Result 4).

4.5.1 Female Workers’ behavior in MNIT

The majority of female participants chose a female avatar in MNIT. In contrast to what the beliefs of female respondents to the post-experimental survey suggest, female workers’ behavior in MNIT was similar to their behavior in EIT rather than MIT. In order to gain some

---

32 In our model, discrimination results from the combination of (i) the firm’s bias in favor of a given gender and (ii) workers’ endogenous avatar choices (i.e., gender reports) in response to this bias.
intuition for possible explanations for our findings, we refer to our theoretical predictions in Section 3. Propositions 1 and 2 suggest that female workers behave differently in MNIT and MIT if the cost of hiding one’s gender and/or the perceived level of discrimination differ depending on whether the task to be performed is known at the avatar choice stage. In particular, in line with our findings, Proposition 1 predicts that the majority of female workers self-identify in MNIT (but not in MIT) if they incur a higher cost from hiding their gender and/or perceive a lower degree of discrimination in MNIT than in MIT (see points (b) and (c) in Proposition 1).

Our post-experimental questionnaire sheds some light on whether differences exist in female workers’ cost to hide their gender and/or their perceived level of discrimination across treatments. We collected information about subjects’ willingness to pay to be perceived as either males or females in the hiring process (see columns 5 and 6 in Panel C of Table 8). When informed about a mathematical task, female participants were willing to pay on average £2.67 more to be perceived as a male rather than a female at the hiring stage. When uninformed about the nature of the task, female participants were willing to pay on average £2.22 more to be perceived as a male rather than a female at the hiring stage. Although these differences are not statistically significant from each other ($t = 0.356, p = 0.745$), it may be that female workers will pay more to hide their gender when they have information about the task to be performed.

Likewise, our survey reveals that female respondents expect a lower percentage of male workers to be hired in MNIT (57.75%) than in MIT (59.87%). The difference is not significant ($t = 0.634, p = 0.53$). Nevertheless, the combination of a lower probability of discrimination perceived by female participants in MNIT versus MIT, together with a higher cost from hiding their gender, may help explain the differences in female workers’ behavior we observe across the two treatments.
4.5.2. Male Workers’ behavior in EIT

The model introduced in Section 3 allows us to offer a simple explanation for male workers’ behavior in EIT. Specifically, the difference between male workers’ behavior in EIT and female workers’ behavior in MIT can be explained by the fact that male workers bear a higher cost from hiding their gender.

As mentioned in Section 4.5.1, we collected data regarding participants’ willingness to pay to be perceived as either males or females in the hiring process. Results show that, when informed about the emotions task, males were willing to pay on average £1.60 extra to be perceived as a female rather than a male at the hiring stage. When informed about a mathematical task, female participants were willing to pay on average £2.67 more to be perceived as a male rather than a female at the hiring stage. This difference is statistically significant ($t = 3.073$, $p = 0.005$, two-tailed). This finding lends support to the idea that males are less willing to hide their gender than females, provided that both workers face a situation in which it might be beneficial to do so.

Evidence from the sociology literature also suggests that males could have a higher cost when switching to a non-male avatar, since male identity is associated with power and effectiveness. In Goldberg (1968), articles written by women or men were given to students for evaluation. They find that articles authored by men received higher ratings. Langford and MacKinnon (2000) find that powerful traits stereotyped as characteristic of men tend to be seen as good while powerless traits stereotyped as characteristic of women tend to be seen as bad.

Subjects in Eagly and Wood (1982) judged that men were more influential and women more easily influenced, even when given no information about the individual’s roles. Eagly, Wood, and Diekman (2000) find that there is a tendency for the specific roles occupied by men
(“providers”) to have more status than the roles occupied by women (“domestic”). Rashotte and Webster test the Ridgeway (1997) view that status beliefs (ideas about competence that are created by gender) arise whenever there are mixed-gender interactions and they find significantly higher general expectations for men than for women.

All in all, it could happen that giving up their perceived status and privileges would lead to males being less willing to choose a non-male avatar.\(^{33}\)

To conclude, in light of our Propositions, our results suggest that only in MIT is anticipated discrimination strong enough to compensate female workers for the relatively small cost of hiding their gender.

4.5.3. Hiring

In our experiment, we find no evidence of discrimination at the hiring stage (Result 4). Our results seem to suggest that there is a difference between “expected” and “actual” discrimination at the hiring stage – i.e., \(\beta_g > \beta_g\), for \(g \in \{x, y\}\) (see also our discussion in Section 4.5.4).

Keeping in mind the limitation of our empirical analysis (see Section 4.2), the absence of discrimination at the hiring stage can be further explained by the presence of a type of employer with a preference for under-represented (unusual) workers, that is, workers’ whose avatar is the scarcest among the pool of available candidates. Results from the post-experimental questionnaire support this assumption: The share of participants who reported a preference for

\[33\] Some readers may feel that these stereotypes are outdated. Yet a recent paper (Charness and Rustichini, 2011) finds strong differences across males and females when they play the Prisoner’s Dilemma in front of audiences: Men are less cooperative when in front of their peers, while women are more cooperative when in front of their peers. These experiments were conducted in California, arguably the forerunner in the women’s movement.
the under-represented gender is 46.47\% in MIT, and 36.67\% in EIT.\textsuperscript{34} By adding this type of employer to the three types introduced in Section 3, equilibria arise in which, even though workers may not have an incentive to decrease the extent to which they hide their gender, discrimination naturally tends to be less prominent at the hiring stage. Appendix I.2 offers a formal analysis in support of these conclusions.

4.5.4. \textit{Further considerations}

In principle, workers might have believed that their avatar choice was meaningless, despite the fact that their experimental instructions stated that firms were not made aware of the avatar-choice stage. If this were to be the case, we would expect workers to either choose their true avatar because of identity considerations, or to randomly choose an avatar. This conclusion does not seem to be supported by our results. We find that 36\% of female workers chose female avatars when informed about the fact that the task was math-related prior to choosing the avatar. By contrast, 64\% of female workers chose female avatars when not informed about the task to be performed. Thus, many female workers seemed to believe that choosing a female avatar would adversely affect the chance of being hired for a math task, which would rule out the idea that workers considered the avatar to be meaningless.

Regarding firms’ behavior, if employers believed that workers chose their own avatar, the gender information conveyed by the avatar might not be meaningful to them. This possibility could be magnified by our design choice of having a neutral avatar. In order to test whether this fact explains the lack of discrimination in our experiment, we conducted an additional treatment that differed from the \textit{MIT} treatment only in that (i) workers could not choose an avatar and were instead exogenously assigned either a neutral avatar or an avatar which reflected their true

\textsuperscript{34} Appendix F also provides evidence in support of this assumption.
gender, and (ii) firms were explicitly told that (and how) the avatars were exogenously assigned to workers. This information was common knowledge to all participants in the experiment.

Our results show that the percentage of male and female avatars hired (as a proportion of all hired avatars) is 48.15% and 44.44%, respectively. Differences are not statistically significant (\(p = 0.699\), test of proportions). So firms behaved no differently than when they were not aware of the avatar-allocation procedure (MIT and MNIT): in neither case do firms discriminate against any of the reported genders. To some extent, this evidence allows us to rule out the lack of trust in the avatar as an explanation for the lack of discrimination we found in the experiment.36

5. Conclusion

This paper considers an employment setting with excess workers, where the only information firms have about a worker is the avatar (male, female, or neutral) chosen by that worker. We examine the choice of the avatar in three treatments. The first involves adding two-digit numbers in which the workers know the task before choosing an avatar, while the second also involves adding numbers; however, in the latter case, workers do not know the nature of the task before selecting avatars. The third treatment involves an emotion-recognition task in a setting where the workers also know the task before picking the avatar.

By looking at the avatar choices, we analyze whether workers react to a potential anticipation of discrimination and how this varies across the type of task and the information condition. Results show that males and females workers react differently when they face a

---

35 For robustness, in the new treatment we use 3 different gender combinations for the groups: (i) 3 male avatars and 2 female avatars, (ii) 2 male and 3 female avatars, and (iii) 2 male avatars, 2 female avatars and 1 neutral avatar.  
36 An alternative explanation for the lack of discrimination is that firms could conceivably be concerned about the experimenter’s view of them. However, given that firms’ payoffs were linked to workers’ performance and that the decisions were completely anonymous, we do not think that this was a very strong driver of firms’ decisions.
situation of potential discrimination. We find that when the task was math-oriented (carrying a male stereotype), most women indeed decided not to reveal their true gender; presumably because they anticipated this would lead to a lower probability of being hired. However, when the task involved emotion recognition, women were much more willing to select an own avatar and did so at the same rate as men. Women also typically chose female avatars when the task was unknown. The difference in the known math-oriented task suggests that women behaved strategically and carefully chose whether to reveal their real gender, depending on the situation.

However, in the emotion-recognition task, which seems to carry a female stereotype, men were quite prone to choose a male avatar regardless. It could be that males were simply not strategic, but our sense is that this is tied to a reluctance to deny the male identity that is considered to be so useful in society. This would mean that a strong sense of power- or privilege-related gender identity (Akerlof and Kranton, 2000) inhibits males from choosing a non-male avatar. Other explanations include overconfidence (Lichtenstein, Fischhoff, and Phillips, 1982; Lundeberg, Fox, and Punccohar, 1994) and differences in emotional reactions to uncertain situations (Croson and Gneezy, 2009; Loewenstein et al., 2001). Nevertheless, more research is needed to clearly identify the main forces driving males’ behavior. Still, the difference in avatar selection by women across treatments indicates responsiveness to the perceived environment.

The patterns in the hiring decisions surprised us. Although there is an expectation in the math task that men would be more likely to be hired than women, given the widespread stereotype that males are better in this sort of tasks, there is no such overall hiring pattern and female avatars are more likely to be hired. Moreover, female avatars are more likely than male avatars to be hired in all three treatments, so discrimination against females was entirely absent.
in our experiment with a standard British subject pool. Our theoretical model provides some support for these findings and reconciles them with the avatar choices.

Regarding performance, we find strong support for the stereotype that men do better on math-related tasks: Men produced 50% more correct sums than women did, and this difference is highly significant. Yet the stereotype that women are better at recognizing emotions received no support in our data, since performance was nearly identical. In fact, we were somewhat surprised by both of these performance results. While there is typically some basis for stereotypes, sometimes they manifest in reality and sometimes they do not.

There has been little previous research regarding how men and women make choices when there is a perception that there could be discrimination in hiring. Anticipation of discrimination is critical in terms of employment-related choices by workers and students intending to become workers. Potentially, our results could have interesting implications for policies aiming to reduce the gender gap in STEM careers, suggesting that it may be even more important to dispel the perception that people have about discrimination than to reduce the discrimination in the labor market, since we do not see evidence of the latter in our data.

In any case, we hope that our findings open a line towards further research on anticipated and actual gender discrimination in different environments.
References


Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. (2016a) “Beliefs about Gender” NBER working paper.


Appendix A. Experimental instructions

A1. Instructions for workers. Part one. Math-Information Treatment

1. There are two different roles assigned in this experiment. A role can be either worker or firm.

2. Your role is worker.

3. Your task for this part will be the following.

   a. You belong to a group that is composed of five workers (including yourself) and two firms.

   b. There will be a market in which the firms have to decide which worker to hire. Each firm will hire only one worker. That means that in your group there will be three workers that will not be hired.

   c. If the firm hires you, the money you and the firm will make will depend on your performance in the following task. You will be adding a series of five two-digit numbers for five minutes. In this case, the piece rate that you and the firm receive per correct sum is £1.5.

   d. If the firm does not hire you, your payoff will still depend on your performance. However, this time the piece rate you receive is £0.5 per correct sum. You do not generate anything for the firm.

   e. The firms do not know anything about you when they are hiring. The only information they will receive will be an avatar that is representing you. The other workers will be also represented by an avatar on the screen.

   f. Your decision will be to choose how you want to be seen by the firms that are going to hire. So, you will have to decide the avatar that represents you in the market. You will be presented with three options and you have to pick the one you prefer.

4. The firms do NOT know that you yourself chose the avatar that represents you and that the other workers themselves also chose the avatar that represents them.

5. After you and the other workers choose your avatars, the firms will see all the workers that they can hire and will decide who to hire.

6. Once everyone is finished, the experimenter will call you using the number of your computer one by one and pay you privately the amount generated in this experiment.
A2. Instructions for firms. Part one. *Math-Information Treatment*

1. There are two different roles assigned in this experiment. A role can be either worker or firm.

2. Your role is firm.

3. Your task for this part will be the following.
   
   a. You belong to a group that is composed of five workers and two firms (including yourself).
   
   b. There will be a market in which the firms have to decide which worker to hire. Each firm will hire only one worker. That means that in your group there will be three workers that will not be hired.
   
   c. The money that you make in this experiment depends on the performance of the worker that you hire.

4. In order to hire, you will be shown a table with all the workers that have been randomly assigned to your group.

5. Once the hiring process is over, the worker you hired will perform a task consisting in a series of sums (five numbers of two digits) during five minutes. You will earn £1.5 per correct sum the worker gets.

6. Once everyone is finished, the experimenter will call you using the number of your computer one by one and pay you privately the amount generated in this experiment.

A3. Instructions for workers who were hired. Part two. *Math-Information Treatment*

1. Congratulations, you have been hired by the firm.

2. For this part of this experiment you will perform a series of sums (five numbers of two digits) during five minutes.

3. You will be paid a piece rate of £1.5 per correct sum you get. In the same way, the firm also receives £1.5 per correct sum you do.
A4. Instructions for workers who were not hired. Part two. Math-Information Treatment

1. You were not hired by the firm.

2. For this part of this experiment you will perform a series of sums (five numbers of two digits) during five minutes.

3. You will be paid a piece rate of £0.5 per correct sum you get. The firm does not make any money from your performance.

A5. Instructions for workers. Part one. Math-No-Information Treatment (the instructions for firms and for the second stage of the experiment were the same as in MIT)

1. There are two different roles assigned in this experiment. A role can be either worker or firm.

2. Your role is worker.

3. Your task for this part will be the following.

   a. You belong to a group that is composed of five workers (including yourself) and two firms.

   b. There will be a market in which the firms have to decide which worker to hire. Each firm will hire only one worker. That means that in your group there will be three workers that will not be hired.

   c. The firms do not know anything about you when they are hiring. The only information they will receive will be an avatar that is representing you. The other workers will be also represented by an avatar on the screen.

   d. Your decision will be to choose how you want to be seen by the firms that are going to hire. So, you will have to decide the avatar that represents you in the market. You will be presented with three options and you have to pick the one you prefer.

4. The firms do NOT know that you yourself chose the avatar that represents you and that the other workers themselves also chose the avatar that represents them.

5. After you and the other workers choose your avatars, the firms will see all the workers that they can hire and will decide who to hire.

6. Later, you will participate in a second stage. If you are hired by the firm, you can make three times more money than if you are not hired. We will explain this with more detail later.
7. Once everyone is finished, the experimenter will call you using the number of your computer one by one and pay you privately the amount generated in the experiment.

A6. Instructions for workers. Part one. Emotions-Information Treatment

1. There are two different roles assigned in this experiment. A role can be either worker or firm.

2. Your role is worker.

3. Your task for this part will be the following.
   a. You belong to a group that is composed of five workers (including yourself) and two firms.
   b. There will be a market in which the firms have to decide which worker to hire. Each firm will hire only one worker. That means that in your group there will be at least three workers that will not be hired.
   c. If the firm hires you, the money you and the firm will make will depend on your performance in the following task. You will be shown 15 photographs depicting individual’s faces. For each image, you will be asked to identify the emotion depicted on the individual’s face. The emotions in the images have been professionally classified by psychologists doing research in this field. The images will be projected on your computer screen for a very short period of time (2 seconds). After the image is shown, you will be given four options from which to select the correctly displayed emotion. You will have 20 seconds to submit your answer. You submit an answer by clicking the submit button with your mouse. You and the firm will get £1.5 per emotion you correctly identify.
   d. If the firm does not hire you, your payoff will still depend on your performance. However, this time the piece rate you receive is £0.5 per emotion you identify. You do not generate anything for the firm.
   e. The firms do not know anything about you when they are hiring. The only information they will receive will be an avatar that is representing you. The other workers will be also represented by an avatar on the screen.
   f. Your decision will be to choose how you want to be seen by the firms that are going to hire. So, you will have to decide the avatar that represents you in the market. You will be presented with three options and you have to pick the one you prefer.
4. The firms do NOT know that you yourself chose the avatar that represents you and that the other workers themselves also chose the avatar that represents them.

5. After you and the other workers choose your avatars, the firms will see all the workers that they can hire and will decide who to hire.

6. Once everyone is finished, the experimenter will call you using the number of your computer one by one and pay you privately the amount generated in this experiment.

A7. Instructions for firms. Part one. Emotions-Information Treatment

1. There are two different roles assigned in this experiment. A role can be either worker or firm.

2. Your role is firm.

3. Your task for this part will be the following.

   a. You belong to a group that is comprised of five workers and two firms (including yourself).

   b. There will be a market in which the firms have to decide which worker to hire. Each firm will hire only one worker. That means that in your group there will be at least three workers that will not be hired.

   c. The money that you make in this experiment depends on the performance of the worker that you hire.

4. In order to hire, you will be shown a table with all the workers that have been randomly assigned to your group.

5. Once the hiring process is over, the worker you hired will perform the following task: workers are shown 15 photographs depicting individual's faces. For each image, they will be asked to identify the emotion depicted on the individual's face. The emotions in the images have been professionally classified by psychologists doing research in this field. The images will be projected on their computer screen for a very short period of time (2 seconds). After the image is shown, they will be given four options from which to select the correctly displayed emotion. They will have 20 seconds to submit the answer. You will get £1.5 per correct emotion the worker identified.
6. Once everyone is finished, the experimenter will call you using the number of your computer one by one and pay you privately the amount generated in this experiment.

A8. **Instructions for workers who were hired. Part two. Emotions-Information Treatment**

1. Congratulations, you have been hired by the firm.

2. For this part of this experiment you will be shown 15 photographs depicting individual's faces. For each image, you will be asked to identify the emotion depicted on the individual's face. The emotions in the images have been professionally classified by psychologists doing research in this field. The images will be projected on your computer screen for a very short period of time (2 seconds). After the image is shown, you will be given four options from which to select the correctly displayed emotion. You will have 20 seconds to submit your answer.

3. You will be paid a piece rate of £1.5 per correct emotion you get. In the same way, the firm also receives £1.5 per correct emotion.

A9. **Instructions for workers who were not hired. Part two. Emotions-Information Treatment**

1. You were not hired by the firm.

2. For this part of this experiment you will be shown 15 photographs depicting individual's faces. For each image, you will be asked to identify the emotion depicted on the individual's face. The emotions in the images have been professionally classified by psychologists doing research in this field. The images will be projected on your computer screen for a very short period of time (2 seconds). After the image is shown, you will be given four options from which to select the correctly displayed emotion. You will have 20 seconds to submit your answer.

3. You will be paid a piece rate of £0.5 per correct emotion you get. The firm does not make any money from your performance.

A10. **Instructions for the boxes experiment.**

You are now taking part in an economic experiment. Depending on your decisions you will be able to earn money. These instructions describe how you can earn money. Please read them carefully.

The money you will make in this experiment will depend on the performance of somebody else on a different task. The task the other person did is as follows. People were shown 15
photographs depicting individual’s faces. For each image, they were asked to identify the emotion depicted on each individual’s face. The emotions in the images have been professionally classified by psychologists doing research in this field. The images were projected on the computer screen for a very short period of time (2 seconds). After the image was shown, participants were given four options of emotions from which they selected the option they thought to be correct. They had 20 seconds to submit their answer.

At the front desk of the lab, you will see two boxes. One of the boxes is labeled “males” and the other one is labeled “females”. Each box contains 20 slips of paper corresponding to 20 individuals who did the task explained above. Each slip has a number printed on it that corresponds to the number of correct emotions identified by an individual who previously did the task.

Your decision involves choosing the box from which to take out one slip of paper. The number that is printed in your paper will determine your payoff. You will get £0.5 times the number of emotions correctly identified by the person you chose.


Your task today will be to guess what other participants did in a previous experiment. Now, we will present different situations that people faced and will ask you a number of questions. Note that the more accurate your guess is, the more money you will earn.

Question 1.

Subjects in a previous experiment participated in the following task. Now, we show the instructions given to these subjects.

1. “For this part of this experiment you will perform a series of sums (five numbers of two digits) during five minutes.
2. You will be paid a piece rate of £1.5 per correct sum you get.”

You have to guess how many correct additions (between 0 and 20) female participants in the previous experiment solved on average. You will earn £5 pounds if your guess is correct. For each correct addition that your guess deviates from the real one, you will lose 50p.

For example, if the real average number of correct answers was 20 and your guess is 18, you will make £5-(2x£0.5)= £4

Number of correct additions for female participants: _____

Question 2.
Subjects in a previous experiment participated in the following task. Now, we show the instructions given to these subjects.

1. “For this part of this experiment you will perform a series of sums (five numbers of two digits) during five minutes.
2. You will be paid a piece rate of £1.5 per correct sum you get.”

You will have to guess how many correct additions (between 0 and 20) male participants in the previous experiment solved on average. You will earn £5 pounds if your guess is correct. For each correct addition that your guess deviates from the real one, you will lose 50p.

For example, if the real average number of correct answers was 20 and your guess is 17, you will make £5-(3x£0.5)= £3.5

Number of correct additions for male participants: _______

Question 3.

Subjects in a previous experiment participated in the following task. Now, we show the instructions given to these subjects.

“Your task for this part will be the following.

1. There are two different roles assigned in this experiment. A role can be either worker or firm.
2. Your role is firm.
   a. There will be a market in which the firms have to decide which worker to hire. Each firm will hire only one worker.
   b. The money that you make in this experiment depends on the performance of the worker that you hire.
3. In order to hire, you will be shown a table with all the workers that have been randomly assigned to your group.
4. Once the hiring process is over, the worker you hired will perform a task consisting in a series of sums (five numbers of two digits) during five minutes. You will earn £1.5 per correct sum the worker gets.”
These are the instructions presented to subjects in the previous experiment. Note that, in this experiment, the firm has information about the workers’ genders. Your task here will be to guess the average percentage of male and female workers hired by the firms. You will earn £5 pounds if your guess is correct. For each percentage point that your guess deviates from the real one, you will lose 50p.

For example, if the percentage of male workers hired is 50% and your guess is 45%, you will make £5-(5x£0.5)= £2.5

Note that the percentages you will guess below must add up to 100%.

Percentage of male workers hired: ______

Percentage of female workers hired: ________

Question 4.

a) In the numbers task explained above, how much would you be willing to pay in order to, when the firm is hiring, being perceived by the firm as a female? Please, choose a number from 0 to 10.

I would be willing to pay _____GBP to be perceived as a female when the firm is hiring.

b) In the numbers task explained above, how much would you be willing to pay in order to, when the firm is hiring, being perceived by the firm as a male? Please, choose a number from 0 to 10.

I would be willing to pay _____GBP to be perceived as a male when the firm is hiring.

Question 5.

Subjects in a previous experiment participated in the following task. Now, we show the instructions given to these subjects.

“Your task for this part will be the following.

1. There are two different roles assigned in this experiment. A role can be either worker or firm.
2. Your role is firm.

   a. There will be a market in which the firms have to decide which worker to hire. Each firm will hire only one worker.

   b. The money that you make in this experiment depends on the performance of the worker that you hire.

3. In order to hire, you will be shown a table with all the workers that have been randomly assigned to your group.

4. Once the hiring process is over, the worker you hired will perform a task consisting in a series of sums (five numbers of two digits) during five minutes. You will earn £1.5 per correct sum the worker gets.”

These are the instructions presented to subjects in the previous experiment. Note that, in this experiment, the firm has information about the workers’ genders. Suppose that, as firm, you face a pool of workers in which one gender (either male or female) is under-represented. Would it be more likely that you hire the worker with an under-represented gender?

Yes ☐  No ☐
### Appendix B. Avatar alternatives

<table>
<thead>
<tr>
<th>Avatar Type</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td><img src="image" alt="Female Avatar" /></td>
</tr>
<tr>
<td>Male</td>
<td><img src="image" alt="Male Avatar" /></td>
</tr>
<tr>
<td>Neutral</td>
<td><img src="image" alt="Neutral Avatar" /></td>
</tr>
</tbody>
</table>
Appendix C. Neutral Avatar choices

Differences in the percentage of men and women choosing the neutral avatar are not statistically significant in any of the treatments ($Z = -0.236, p = 0.813; Z = 1.427, p = 0.153; \text{ and } Z = 1.427, p = 0.153$, two-tailed Mann-Whitney test for MIT, MNIT and EIT, respectively).
Appendix D. Avatar hiring rates

The approach to study hiring decisions provided in the main text does not take into account the supply of avatars that the firms face. Another methodology defines hiring rates in terms of the proportion hired from the available supply of avatars. Figure D1 plots the probability that a worker who chose a particular avatar (male or female) is hired. We do that by computing, from all the male and female avatars that were chosen by the workers, the proportion of them that were hired in the labor market.

As observed in Figure D1, in the MIT treatment, the hiring rate for male avatars is 27.78% (there are 54 male avatars, of which 15 are hired); in MNIT, this hiring rate is 35.56% (16 of 45). There is no significant difference between these rates ($Z = 0.831, p = 0.406$, test of proportions). The hiring rate for female avatars are 60.87% (14 of 23) in MIT, and 37.78% (17 of 45) in MNIT. The difference between these rates is marginally significant ($Z = 1.809, p = 0.071$, test of proportions). Note that firms have the same information in both cases, so that there is an argument for pooling these data in any case. If we do so, the conditional hiring rates are 31.3% (31 of 99) for male avatars and 45.6% (31 of 68) for female avatars; this difference is marginally significant ($Z = 1.876, p = 0.061$, test of proportions).
Figure D1: Probability of being hired in each treatment, by avatar

![Bar chart showing probability of being hired by avatar and treatment.](image)

Turning to the emotion-recognition task, the hiring rates are 32.65% (16 out of 49) for male avatars and 51.22% (21 out of 41) for the female avatars. This difference is marginally significant ($Z = 1.783, p = 0.075$, test of proportions).

Overall, this analysis does not show any hiring discrimination against female avatars. To the extent that there are differences, we find marginally-significant discrimination in favor of female avatars in both the adding-numbers and emotions-recognition tasks. In fact, if we pool the data from all three treatments, we see that 33.78% (50 of 148) of male avatars were hired and 47.71% (52 of 109) female avatars were hired; this difference is statistically significant ($Z = 2.255, p = 0.024$). So, female avatars seemed to be more attractive hires than male avatars.

A third form of analysis considers another way that the distribution of avatars can affect hiring patterns. The lower proportion of male avatars hired by firms reported earlier in this section could reflect two different issues. While it could be the case that female avatars are *per se* more attractive to firms than male avatars, it could also be that there are more male avatars chosen than those that could be hired (recall that only two of five avatars could be hired),
introducing some sort of bias in our previous results.\textsuperscript{37} In order to control for this, we calculate, for each combination of avatars, the number of male and female avatars that were hired as a fraction of the maximum number that could be hired.\textsuperscript{38} Figure D2 plots the fraction of avatars (male or female) that are hired, controlling for the particular composition of the group.

Pooling data from \textit{MIT} and \textit{MNIT}, the fraction of male and female avatars hired by companies, as a percentage of the maximum number of avatars that could be hired given the composition of the group, was 41\%, and 55\%, respectively.\textsuperscript{39} In \textit{EIT}, the respective figures are 38\% and 60\% for male and female avatars, respectively.\textsuperscript{40} As with the previous analysis, the results indicate that there is positive discrimination towards female avatars.

\begin{flushright}
$^\text{37}$In fact, such an overabundance of one avatar type did occur. In \textit{MIT}, there are 10 cases in which there were more than two male avatars available, and two cases with more than two female avatars. In \textit{MNIT}, there are eight cases in which there were more than two male avatars available, and seven cases with more than two female avatars. In \textit{EIT}, there are nine cases in which there were more than two male avatars available, and five cases with more than two female avatars. Note that there are more cases of more than two male avatars than more than two female avatars.

$^\text{38}$For example, consider a group composed of three male avatars, one female avatar and one neutral avatar. The maximum number of male avatars that could be hired in that group would be two (assuming that both firms hire a male avatar), the maximum number of female avatars that could be hired in that group is one, the same as the maximum number of neutral avatars. If in this case, for example, one male avatar and one neutral avatar were hired, we would say that the fraction of male avatars hired in that group is 50\% and the fraction of female avatars hired in that group is 100\%.

$^\text{39}$Differences are statistically significant ($p = 0.048$, paired $t$-test).

$^\text{40}$Differences are statistically significant ($p < 0.001$, paired $t$-test).
Figure D2: Fraction of avatars hired, by treatment

<table>
<thead>
<tr>
<th>Male Avatar</th>
<th>Female Avatar</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT</td>
<td>36.68</td>
</tr>
<tr>
<td>MNIT</td>
<td>44.90</td>
</tr>
<tr>
<td>EIT</td>
<td>38.00</td>
</tr>
</tbody>
</table>
Appendix E. Neutral avatar hiring rates, by treatment
Appendix F. Analysis of under-represented avatars

In the model presented in Section 3.5, type 1 \( P \) hires the worker whose reported gender is underrepresented. In this section, we check whether the assumption regarding the presence of such a type of firm is supported by our data. We observe that relative scarcity seems to be a fairly good predictor of hiring rates. If we regress the hiring-rate data in Figure D1 in Appendix D and Appendix E (nine observations, one for each type of avatar in each treatment) against the proportion of each avatar chosen in the respective treatment, we have this simple regression:

\[
\text{Hiring rate of avatar} = 60.07 - 51.21 \times \text{Avatar proportion} \quad \text{Adj. } R^2 = 0.291
\]

\begin{align*}
\text{Hiring rate of avatar} & = 60.07 - 51.21 \times \text{Avatar proportion} \\
& \quad (8.94) \quad (24.74)
\end{align*}

This gives \( t = -2.07 \) for the coefficient of the avatar proportion, significant at \( p = 0.077 \). Thus, there may be a preference for less-frequent avatars, as if firms have a preference for unusual workers.

\[41\text{ An alternative regression using the hiring-rate data from the analysis in Figure 5 and Appendix C gives similar results, with } t = -1.68 \text{ for the coefficient of the avatar proportion, significant at } p = 0.136. \text{ The significance levels are low for the } t\text{-statistics because there are only nine observations.}\]
Appendix G. Stereotype of the emotions task

To test whether the emotions task carries a female stereotype, we conducted an additional experiment with 40 subjects (20 males and 20 females). Similarly to Aguiar et al (2009), two different boxes labeled “female” and “male” were placed at the front of a room. Each box contained 20 slips of paper. Each slip was printed with the number of correct emotions identified by the workers in the experiment. 42

The decision was simply to choose the box from which to select one slip of paper. The number printed in the paper would determine subjects’ payoffs. Participants were paid 0.5 GBP times the number on the slip of paper selected. This was common information. Subjects approached the boxes one by one and made their decision privately. Once the decision was made, they would show the slip of paper to the experimenter and would receive their money. The slip of paper was then put back in the corresponding box.

We find that 85% of the population took the slip of paper from the “female” box, showing that people believe that women are better than men at this task; the binomial test on the entire population gives $p = 0.000$. When we distinguish by gender, results are very similar for males and females. Eighty percent of male ($p = 0.007$) and 90% of female participants ($p = 0.000$) picked the slip of paper from the “female” box.

This finding is confirmed by our post-experimental survey: Respondents believed that female participants in the experiment performed significantly better than male participants at recognizing emotions (see Appendix J). All in all, the emotions task was heavily considered to

42 The 40 workers reported in the slips of paper were randomly selected from the entire subject pool that had participated in the main experiment.
be a female-oriented task. We can therefore rule out the explanation that the stereotype of this task was not prevalent.
Appendix H. Proofs of Propositions

Proof of Proposition 1. Suppose $\beta_y > \beta_x = 0$. $A_2$'s ex-ante utility from misreporting her/his gender is:

$$U_2(y,x) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] + \beta_y \left[ (s(1-p) + p(1-s) + (1-p)(1-s)) \bar{u} + ps \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] + \beta_x \left[ (s(1-p) + p(1-s)) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + sp \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + (1-p)(1-s) \bar{u} - C_y \right].$$

(5)

whereas $A_2$'s (ex-ante) utility from reporting $g_2$, i.e., $r_2 = y$, is:

$$U_2(y,y) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] + \beta_y \left[ (s(1-p) + p(1-s)) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + (1-p)(1-s) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + ps \bar{u} \right] + \beta_x \left[ (s(1-p) + p(1-s) + sp) \bar{u} + (1-p)(1-s) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right].$$

(6)

From (5) and (6), $A_2$'s dominant strategy consists in truthfully reporting $r_2 = y$, i.e., $q^* = 0$. Further, $A_3$'s ex-ante utility from truthfully reporting her/his gender is:
\[ U_3(x, x) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]

\[ + \beta_y \left[ (q(1 - p) + p(1 - q) + (1 - p)(1 - q))u + ps \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] \]

\[ + \beta_x \left[ (q(1 - p) + p(1 - q)) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + qp \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] \]

\[ + (1 - p)(1 - q) \bar{u} \]. \quad (7) \]

whereas \( A_3 \)'s (ex-ante) utility from misreporting \( g_3 \) is:

\[ U_3(x, y) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]

\[ + \beta_y \left[ (q(1 - p) + p(1 - q)) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + (1 - p)(1 - q) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + pq \bar{u} \right] \]

\[ + \beta_x \left[ (q(1 - p) + p(1 - q) + qp)u + (1 - p)(1 - q) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] \]

\[ - c_x. \quad (8) \]

From (2)-(3) and (7)-(8), if \( \frac{c_x}{\beta_y \Delta u} < \frac{1}{3 \cdot 2} \), both \( A_1 \) and \( A_3 \) have a dominant strategy, which consists in misreporting their gender, i.e., \( p^* = s^* = 0 \). This proves point (a) in the Proposition.

If \( \frac{c_x}{\beta_y \Delta u} \geq \frac{1}{2} \), \( A_1 \) and \( A_3 \) have a dominant strategy, which consists in truthfully reporting their gender, i.e., \( p^* = s^* = 1 \). This proves point (c) in the Proposition. Further, if \( \frac{c_x}{\beta_y \Delta u} \in \left( \frac{1}{3 \cdot 2}, \frac{1}{2} \right) \), from (2)-(3) and (7)-(8), three equilibria exist, namely \((p^*, q^*, s^*) = \{ (0,0,1), (1,0,0), \left( 2 \left[ 3 \frac{c_x}{\beta_y \Delta u} - 1 \right] \right) \)\}. This proves point (b) in the Proposition.\]
**Proof of Proposition 2.** See Proof of Proposition 1.\[\]  

**Proof of Proposition 3.** Suppose $\hat{\beta}_x > \hat{\beta}_y = 0$. From (2)-(3) and (7)-(8), $A_1$ and $A_3$ have a dominant strategy, which consists in truthfully reporting $r_1 = r_3 = x$, that is, $p^* = s^* = 1$. From (5) and (6), $A_2$ plays $r_2 = x$ and $q^* = 1$ if $\frac{c_x}{\beta_y u} < \frac{1}{3}$. This proves point (a) in the Proposition. $A_2$ plays $r_2 = y$ and $q^* = 0$ otherwise. This proves point (b) in the Proposition.\[\]
Appendix I. Extensions of the model

In this Appendix, we consider two extensions of the model and discuss their implications for the equilibria described in Section 3. In Appendix I.1, we introduce the possibility for workers to hide their gender by either reporting a “neutral gender” or the “opposite” gender. In Appendix I.2, we introduce a type of firm with preference for workers characterized by an under-represented gender, that is workers’ whose reported gender is the scarcest among the pool of available candidates.

I.1. Neutral gender

Let us consider a model in which each worker $A_i$ can make a report $r_i \in \{x, n, y\}$, for $i = 1, 2, 3$, where $n$ corresponds to the neutral avatar in our experiment, and where $g_i \in \{x, y\}$. We assume that, when facing $r_i = n$, $P$ attaches an equal probability to $g_i = x$ and $g_i = y$. In particular, a firm of type $\theta_1$ prefers to hire a worker whose report is $y$ over a worker whose report is either $n$ or $x$. Also, type $\theta_1$ prefers to hire a worker whose report is $n$ over a worker whose report is $x$. Similarly, a firm of type $\theta_2$ prefers to hire a worker whose report is $x$ over a worker whose report is either $n$ or $y$. Also, type $\theta_2$ prefers to hire a worker whose report is $n$ over a worker whose report is $y$.

Let $C_{g_i}^{r_i}$ denote the cost for $A_i$ of making a report $r_i \in \{g_i, n, -g_i\}$, for $g_i, -g_i \in \{x, y\}$ and $g_i \neq -g_i$, with $C_{g_i}^{g_i} = 0$. In particular, $C_{x}^{y}$ and $C_{x}^{n}$ denote the cost for a worker whose gender is $x$ to report $y$ and $n$, respectively. In line with our reasoning in footnote 9, we assume $C_{x}^{y} \geq C_{x}^{n} \geq 0$. Similarly, let $C_{y}^{x}$ and $C_{y}^{n}$ denote the cost for a worker whose gender is $y$ to report $x$ and
Let $p_x$ and $p_n$ denote the probability that $A_1$ reports $r_1 = x$ and $r_1 = n$, respectively. Likewise, we denote by $q_x$ and $q_n$ the probability that $A_2$ reports $r_2 = x$ and $r_2 = n$, respectively. Finally, we denote by $s_x$ and $s_n$ the probability that $A_3$ reports $r_3 = x$ and $r_3 = n$, respectively.

$A_1$’s ex-ante utility from truthfully reporting $r_1 = x$ is:

$$U_1(x, x) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \overline{u} + \frac{2}{3} u \right]$$

$$+ \beta_y \left[ (1 - q_x - q_n)(1 - s_x - s_n)\overline{u} + q_n(1 - s_x - s_n)u + s_n(1 - q_x - q_n)u \right]$$

$$+ q_n s_n \overline{u} + q_x (1 - s_x - s_n)u + s_x (1 - q_x - q_n)u + q_x s_n u + q_n s_x u$$

$$+ q_x s_x \left( \frac{1}{3} \overline{u} + \frac{2}{3} u \right) \right]$$

$$+ \beta_x \left[ (1 - q_x - q_n)(1 - s_x - s_n)\overline{u} + q_n(1 - s_x - s_n)\overline{u} + s_n(1 - q_x - q_n)\overline{u} \right]$$

$$+ q_n s_n \overline{u} + q_x (1 - s_x - s_n) \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right) + s_x (1 - q_x - q_n) \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right)$$

$$+ q_x s_n \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right) + q_n s_x \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right) + q_x s_x \left( \frac{1}{3} \overline{u} + \frac{2}{3} u \right) \right] \right]$$

$A_1$’s ex-ante utility from reporting $r_1 = n$ is:

---

43 When needed, we assume that $A_i$ truthfully reports her/his gender when she/he is indifferent between making a truthful report ($r_i = g_i$) and hiding her/his gender ($r_i \in \{n, -g_i\}$), for $i = 1, 2, 3$. Similarly, we assume that $A_i$ reports $r_i = -g_i$ when she/he is indifferent between $r_i = -g_i$ and $r_i = n$. 
\[ U_1(x, n) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \\
+ \beta_y \left[ (1 - q_x - q_n)(1 - s_x - s_n) \bar{u} + q_n (1 - s_x - s_n) u + s_n (1 - q_x - q_n) u \right] \\
+ q_n s_n \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + q_x (1 - s_x - s_n) \bar{u} + s_x (1 - q_x - q_n) u + q_x s_n \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \\
+ q_n s_x \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_x s_x \bar{u} \\
+ \beta_x \left[ (1 - q_x - q_n)(1 - s_x - s_n) \bar{u} + q_n (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \\
+ s_n (1 - q_x - q_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_n s_n \bar{u} + q_x (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \\
+ s_x (1 - q_x - q_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_x s_n \bar{u} + q_n s_x \bar{u} + q_x s_x \bar{u} \right] \\
+ s_n (1 - q_x - q_n) u + q_n s_n u + q_x (1 - s_x - s_n) u + s_x (1 - q_x - q_n) u + q_x s_n u \\
+ q_n s_x u + q_x s_x \bar{u} \right] - C_x^n. \quad (10) \]

\[ U_1(x, y) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \\
+ \beta_y \left[ (1 - q_x - q_n)(1 - s_x - s_n) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + q_n (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \\
+ s_n (1 - q_x - q_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_n s_n \bar{u} + q_x (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \\
+ s_x (1 - q_x - q_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_x s_n \bar{u} + q_n s_x \bar{u} + q_x s_x \bar{u} \right] \\
+ \beta_x \left[ (1 - q_x - q_n)(1 - s_x - s_n) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + q_n (1 - s_x - s_n) u \\
+ s_n (1 - q_x - q_n) u + q_n s_n u + q_x (1 - s_x - s_n) u + s_x (1 - q_x - q_n) u + q_x s_n u \\
+ q_n s_x u + q_x s_x \bar{u} \right] - C_x^y. \quad (11) \]
For the sake of completeness, we also report $A_2$’s and $A_3$’s utility functions. $A_2$’s ex-ante utility from reporting the opposite gender, i.e., $r_2 = x$, is:

$$U_2(y, x) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \overline{u} + \frac{2}{3} u \right]$$

$$+ \beta_y \left[ (1 - p_x - p_n)(1 - s_x - s_n)\overline{u} + p_n (1 - s_x - s_n)u + s_n (1 - p_x - p_n)\overline{u} \right.$$

$$+ p_n s_n u + p_x (1 - s_x - s_n)u + s_x (1 - p_x - p_n)\overline{u} + p_x s_n u + p_n s_x u$$

$$+ p_x s_x \left[ \frac{1}{3} \overline{u} + \frac{2}{3} u \right] \right]$$

$$+ \beta_x \left[ (1 - p_x - p_n)(1 - s_x - s_n)\overline{u} + p_n (1 - s_x - s_n)\overline{u} + s_n (1 - p_x - p_n)\overline{u} \right.$$

$$+ p_n s_n \overline{u} + p_x (1 - s_x - s_n) \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right) + s_x (1 - p_x - p_n) \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right)$$

$$+ p_x s_n \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right) + p_n s_x \left( \frac{1}{2} \overline{u} + \frac{1}{2} u \right) + p x s_x \left( \frac{1}{3} \overline{u} + \frac{2}{3} u \right) \right] - C_{y} \chi. \quad (12)$$

$A_2$’s ex-ante utility from reporting $r_2 = n$ is:
\[ U_2(y, n) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]

\[ + \beta_y \left[ (1 - p_x - p_n) (1 - s_x - s_n) u + p_n (1 - s_x - s_n) u + s_n (1 - p_x - p_n) u \right] \]

\[ + p_n s_n \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + p_x (1 - s_x - s_n) u + s_x (1 - p_x - p_n) u + p_x s_n \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \]

\[ + p_n s_x \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + p_x s_x \bar{u} \]

\[ + \beta_x \left[ (1 - p_x - p_n) (1 - s_x - s_n) \bar{u} + p_n (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \right] \]

\[ + s_n (1 - p_x - p_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + p_n s_n \bar{u} + p_x (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \]

\[ + s_x (1 - p_x - p_n) \bar{u} + p_x s_n u + p_n s_x u + p_x s_x \bar{u} \]

\[ + \beta_x \left[ (1 - p_x - p_n) (1 - s_x - s_n) \bar{u} + p_n (1 - s_x - s_n) u \right] \]

\[ + s_n (1 - p_x - p_n) u + p_n s_n u + p_x (1 - s_x - s_n) u + s_x (1 - p_x - p_n) u + p_x s_n u \]

\[ + p_n s_x u + p_x s_x u \]. \quad (13) \]

\[ A_2 \text{'s ex-ante utility from truthfully reporting } r_2 = y \text{ is:} \]

\[ U_2(y, y) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]

\[ + \beta_y \left[ (1 - p_x - p_n) (1 - s_x - s_n) \bar{u} + p_n (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \right] \]

\[ + s_n (1 - p_x - p_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + p_n s_n \bar{u} + p_x (1 - s_x - s_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \]

\[ + s_x (1 - p_x - p_n) \bar{u} + p_x s_n u + p_n s_x u + p_x s_x \bar{u} \]

\[ + \beta_x \left[ (1 - p_x - p_n) (1 - s_x - s_n) \bar{u} + p_n (1 - s_x - s_n) u \right] \]

\[ + s_n (1 - p_x - p_n) u + p_n s_n u + p_x (1 - s_x - s_n) u + s_x (1 - p_x - p_n) u + p_x s_n u \]

\[ + p_n s_x u + p_x s_x u \]. \quad (14) \]
The ex-ante utility from truthfully reporting $r_3 = x$ is:

\[
U_3(x, x) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right]
\]

\[
+ \beta_y \left[ (1 - q_x - q_n)(1 - p_x - p_n)u + q_n(1 - p_x - p_n)\bar{u} + p_n(1 - q_x - q_n)\bar{u} + q_n p_n u + q_x (1 - p_x - p_n) u + p_x (1 - q_x - q_n) u + q_x p_n u + q_n p_x u \right]
\]

\[
+ q_x p_x \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right]
\]

\[
+ \beta_x \left[ (1 - q_x - q_n)(1 - p_x - p_n)\bar{u} + q_n(1 - p_x - p_n)\bar{u} + p_n(1 - q_x - q_n)\bar{u} + q_n p_n \bar{u} + q_x (1 - p_x - p_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + p_x (1 - q_x - q_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \right]
\]

\[
+ q_x p_n \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_n p_x \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_x p_x \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right].
\]

\[ (15) \]

The ex-ante utility from reporting $r_3 = n$ is:
\[ U_3(x, n) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]

\[ + \beta_y \left[ (1 - q_x - q_n)(1 - p_x - p_n)u + q_n (1 - p_x - p_n)u + p_n (1 - q_x - q_n)u \right. \]

\[ + q_n p_n \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + q_x (1 - p_x - p_n)u + p_x (1 - q_x - q_n)u \]

\[ + q_x p_n \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_n p_x \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_x p_x \bar{u} \]

\[ + \beta_x \left[ (1 - q_x - q_n)(1 - p_x - p_n)\bar{u} + q_n (1 - p_x - p_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \right. \]

\[ + p_n (1 - q_x - q_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_n p_n \bar{u} + q_x (1 - p_x - p_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \]

\[ + p_x (1 - q_x - q_n) u + q_x p_n u + q_n p_x u + q_x p_x u \right] - C_x^u. \]  \(16\)

Finally, \(A_3\)'s ex-ante utility from reporting the opposite gender, i.e., \(r_3 = y\), is:

\[ U_3(x, y) = (1 - \beta_y - \beta_x) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]

\[ + \beta_y \left[ (1 - q_x - q_n)(1 - p_x - p_n) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + q_n (1 - p_x - p_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \right. \]

\[ + p_n (1 - q_x - q_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + q_n p_n \bar{u} + q_x (1 - p_x - p_n) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) \]

\[ + p_x (1 - q_x - q_n) u + q_x p_n u + q_n p_x u + q_x p_x u \right] \]

\[ + \beta_x \left[ (1 - q_x - q_n)(1 - p_x - p_n) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + q_n (1 - p_x - p_n) u \right. \]

\[ + p_n (1 - q_x - q_n) u + q_n p_n u + q_x (1 - p_x - p_n) u + p_x (1 - q_x - q_n) u \]

\[ + q_x p_n u + q_n p_x u + q_x p_x u \right] - C_x^y. \]  \(17\)
We can now state the following result:

**Lemma 1.** At equilibrium, $p_n^* = 0$ if $1 - q_n^* - q_x^* = 1$ and/or $1 - s_n^* - s_x^* = 1$. Likewise, $s_n^* = 0$ if $1 - q_n^* - q_x^* = 1$ and/or $1 - p_n^* - p_x^* = 1$. Finally, $q_n^* = 0$ if $p_n^* = 1$ and/or $s_n^* = 1$.

**Proof of Lemma 1:** Consider a putative equilibrium in which $A_2$ and/or $A_3$ report $y$ with probability 1. Then, from (9)-(10), we have $U_1(x, n) \leq U_1(x, x)$, which implies $p_n^* = 1$. Similarly, consider a putative equilibrium in which $A_1$ and/or $A_2$ reports $y$ with probability 1. Then, from (15)-(16), we have $U_3(x, n) \leq U_3(x, x)$, which implies $s_n^* = 1$. Finally, consider a putative equilibrium in which $A_1$ and/or $A_3$ report $x$ with probability 1. Then, from (13)-(14), we have $U_2(y, n) \leq U_2(y, y)$, which implies $q_n^* = 0$. 

According to Lemma 1, each worker prefers to make a truthful report $r_i = g_i$ rather than reporting a neutral gender whenever she/he attaches probability 1 to at least one worker reporting $g_{-i}$, for $g_i, g_{-i} \in \{x, y\}$ and $g_i \neq g_{-i}$, and $i = 1, 2, 3$. Intuitively, compared to making a truthful report, reporting the neutral gender involves both costs – i.e., $C_n^{g_i}$ – and benefits – i.e., it allows $A_i$ to potentially avoid discrimination by the type of $P$ who prefers to hire workers who report $g_{-i}$. The benefits however materialize only if none of the other workers reports $r_j = g_{-i}$, for $i, j = 1, 2, 3$ and $i \neq j$.

We can now state the following Proposition:

**Proposition 4.** Propositions 1, 2 and 3 continue to hold in a setting in which $A_i$ can make a report $r_i \in \{x, n, y\}$, for $i = 1, 2, 3$. 

68
Proof of Proposition 4: Suppose $\beta_y > \beta_x = 0$. From (12)-(13)-(14), $r_2 = y$ is $A_2$’s weakly dominant strategy. Hence, from Lemma 1, neither $A_1$ nor $A_3$ find it profitable to report the neutral gender. The result then follows from the proof of Proposition 2. Similarly, when $\beta_x > \beta_y = 0$, from (9)-(10)-(11) and (15)-(16)-(17), $A_1$’s and $A_3$’s weakly dominant strategies are $r_1 = r_3 = x$. Hence, from Lemma 1, $A_2$ does not find it profitable to report the neutral gender. The result then follows from the proof of Proposition 3.

Proposition 4 confirms our main findings in Section 3. According to this Proposition, in our simple setting, reporting a neutral gender does not occur at equilibrium. Intuitively, when workers attach a zero value to either $\beta_y$ or $\beta_x$, the worker whose gender is not discriminated does not have an incentive to hide her/his gender. This, in turn, implies that, for workers whose gender is discriminated, the benefit from reporting a neutral gender rather than their real gender is null (see Lemma 1).

To gain further intuition on the role played by the neutral gender, we briefly discuss the case in which both $\hat{\beta}_y > 0$ and $\hat{\beta}_x > 0$. It is important to note that allowing for $\hat{\beta}_y$ and $\hat{\beta}_x$ to simultaneously take strictly positive values does not affect the results reported in Propositions 1, 2 and 3 in any major way. Yet, provided that $\hat{\beta}_y$ and $\hat{\beta}_x$ are not too different and that hiding one’s gender is not too costly, making a truthful report is not a dominant strategy for some of the workers any longer. As a consequence, reporting a neutral gender can become part of the equilibrium play. To better understand the trade-offs involved in workers’ choices, consider $A_1$, with $g_1 = x$. By inspection of (9) and (10), $\forall \{r_2, r_3\}$, reporting $r_1 = n$ rather than $r_1 = x$ entails

---

44 See footnote 22.
45 Given the large number of unknowns, equations, and parameters involved in the analysis, we do not attempt to compute the equilibria of the game, as this analysis turns out to be computationally cumbersome.
both costs – i.e., $C^n_2$ and a (weakly) lower probability of being hired by a firm of type $\theta_2$ – and benefits – i.e., it (weakly) increases the odds of being hired by a firm of type $\theta_1$. By inspection of (10) and (11), $\forall \{r_2, r_3\}$, reporting $r_1 = n$ rather than $r_1 = y$ entails both costs – a (weakly) lower probability of being hired by a firm of type $\theta_1$ – and benefits – a lower cost of hiding her\'s gender and a (weakly) higher probability of being hired by a firm of type $\theta_2$.

Overall, hiding one’s gender by reporting a neutral gender rather than the opposite gender allows a worker to hedge her\'s position when both genders face discrimination with a strictly positive probability. Yet, this reasoning does not invalidate the main message of the paper: Workers have an incentive to hide their gender – by reporting either the neutral gender or the opposite gender – when they anticipate discrimination.

I.2. Under-represented gender

Let us consider a model in which $\theta \in \{\theta_1, \theta_2, \theta_3, \theta_4\}$. The types $\theta_1, \theta_2,$ and $\theta_3$ are defined as in Section 3. A firm $P$ of type $\theta_4$ hires a worker whose reported gender is under-represented in the pool of available reports. Type $\theta_4$ hires each worker with equal probability if all workers report the same gender. We denote by $\beta_u$ the probability that $\theta = \theta_4$. Because the frequency of this type of firm is not related to the task $t$ to be performed by workers, we set $\beta_u = \hat{\beta}_u$, for $t \in \{t_x, t_y\}$ and independently of workers’ information about the task, with $\hat{\beta}_u$ denoting workers’ expected probability that $\theta = \theta_4$.46

For the sake of simplicity, we revert to the model in which $r_i \in \{x, y\}$, for $i = 1,2,3$. $A_1$’s (ex-ante) utility from truthfully reporting her\'s gender is:

46 Results from the post-experimental survey show that the fraction of subjects who reported to have a preference for under-represented genders is 46.47% in MIT, and 36.67% in EIT (see Table 8, Panel A, in Appendix J). Differences across treatments are not statistically significant ($t = 0.776, p = 0.441$, two-tailed t-test).
\[ U_1(x, x) = (1 - \beta_y - \beta_x - \beta_u) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]
\[ + \beta_y \left[ (s(1 - q) + q(1 - s) + (1 - q)(1 - s))\bar{u} + qs \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] \]
\[ + \beta_x \left[ (s(1 - q) + q(1 - s)) \left( \frac{1}{2} \bar{u} + \frac{1}{2} u \right) + sq \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + (1 - q)(1 - s)\bar{u} \right] \]
\[ + \beta_u \left[ (1 - q)(1 - s)\bar{u} + (q(1 - s) + s(1 - q))\bar{u} + qs \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] , \quad (18) \]

whereas \( A_1 \)'s (ex-ante) utility from (mis)reporting \( g_1 \), i.e., \( r_1 = y \), is:
\[ U_1(x, y) = (1 - \beta_y - \beta_x - \beta_u) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]
\[ + \beta_y \left[ (s(1 - q) + q(1 - s))(\frac{1}{2} \bar{u} + \frac{1}{2} u) + (1 - q)(1 - s) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + qs\bar{u} \right] \]
\[ + \beta_x \left[ (s(1 - q) + q(1 - s) + sq)u + (1 - q)(1 - s) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] \]
\[ + \beta_u \left[ (1 - q)(1 - s) \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + (q(1 - s) + s(1 - q))\bar{u} + qs\bar{u} \right] - C_x. \quad (19) \]

\( A_3 \)'s utilities are similarly defined, where we substitute \( p \) for \( s \). \( A_2 \)'s (ex-ante) utility from misreporting her/his gender is:
\[ U_2(y, x) = (1 - \beta_y - \beta_x - \beta_u) \left[ \frac{1}{3} \bar{u} + \frac{2}{3} u \right] \]
\[ + \beta_y \left[ (s(1 - p) + p(1 - s) + (1 - p)(1 - s))\bar{u} + ps \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] \]
\[ + \beta_x \left[ (s(1 - p) + p(1 - s))(\frac{1}{2} \bar{u} + \frac{1}{2} u) + pq \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) + (1 - p)(1 - s)\bar{u} \right] \]
\[ + \beta_u \left[ (1 - p)(1 - s)\bar{u} + (p(1 - s) + s(1 - p))\bar{u} + ps \left( \frac{1}{3} \bar{u} + \frac{2}{3} u \right) \right] - C_y, \quad (20) \]
whereas $A_2$’s (ex-ante) utility from truthfully reporting $g_2$, i.e., $r_2 = y$, is:

\[
U_2(y, y) = (1 - \beta_y - \beta_x - \beta_u) \left[ \frac{1}{3} \tilde{u} + \frac{2}{3} u \right] + \beta_y \left[ (s(1 - p) + p(1 - s)) \left( \frac{1}{2} \tilde{u} + \frac{1}{2} u \right) + (1 - p)(1 - s) \left( \frac{1}{3} \tilde{u} + \frac{2}{3} u \right) + ps\tilde{u} \right] + \beta_x \left[ (s(1 - p) + p(1 - s) + sp)u + (1 - p)(1 - s) \left( \frac{1}{3} \tilde{u} + \frac{2}{3} u \right) \right] + \beta_u \left[ (1 - p)(1 - s) \left( \frac{1}{3} \tilde{u} + \frac{2}{3} u \right) + (p(1 - s) + s(1 - p))u + ps\tilde{u} \right]. 
\] (21)

We also define the total probability that a worker whose reported gender is $x$ is hired:

\[
H_x = pqs + [qs(1 - p) + pq(1 - s) + ps(1 - q)] \left[ \beta_x + \frac{2}{3} (1 - \beta_y - \beta_x - \beta_u) \right] + [s(1 - p)(1 - q) + q(1 - p)(1 - s) + p(1 - q)(1 - s)] \left[ \beta_x + \beta_u + \frac{1}{3} (1 - \beta_y - \beta_x - \beta_u) \right]. 
\] (22)

In what follows, we focus on the following case: $\beta_u > 0$ and $\beta_y > \beta_x = 0$, with $\beta_u > 0$ and $\beta_y > \beta_x = 0$. In general, compared to Propositions 1 and 2, the presence of the type $\theta_4$ may prevent $r_2 = y$ from being a dominant strategy for $A_2$.\footnote{In particular, $A_2$ has a dominant strategy ($r_2 = y$) if $\frac{C_x}{\beta_u\Delta u} \geq \max \left\{ \frac{2p_2}{3\beta_y} - \frac{1}{3}, 0 \right\}$.} Formally, from (18)-(19) and (20)-(21) we have:

\[
U_1(x, x) - U_1(x, y) \iff -\frac{\beta_x}{3} - \frac{q + s}{6} \frac{\beta_x}{3} + \frac{2}{3} \beta_u(1 - q - s) + \frac{C_x}{\Delta u} \leq 0, 
\] (23)
\[ U_2(y,x) - U_2(y,y) \iff -\frac{\beta_x}{3} - \frac{p+s}{6} - \frac{2}{3} \beta_u (1 - p - s) - \frac{C_y}{\Delta u} \leq 0, \quad (24) \]

\[ U_3(x,x) - U_3(x,y) \iff -\frac{\beta_x}{3} - \frac{p+q}{6} - \frac{2}{3} \beta_u (1 - p - q) + \frac{C_x}{\Delta u} \leq 0. \quad (25) \]

We can state the following results:

**Proposition 5.** Suppose \( \beta_u > 0 \) and \( \beta_y > \beta_x = 0 \), with \( \beta_u > 0 \) and \( \beta_y > \beta_x = 0 \). Then,

- **a.** if \( \frac{c_y}{\beta_y \Delta u} < \frac{1}{3} - \frac{2}{3} \beta_u \beta_y \), for \( g \in \{x, y\} \), there exists an equilibrium in which \( A_1 \) and \( A_3 \) misreport their gender, whereas \( A_2 \) makes a truthful report, that is, \( \{p^*, q^*, s^*\} = \{0,0,0\} \);
- **b.** if \( \frac{c_x}{\beta_x \Delta u} \in \left[ \frac{1}{3} - \frac{2}{3} \beta_u \beta_y - \frac{1}{2} \right] \), there exist two equilibria in which one between \( A_1 \) and \( A_3 \) makes a non-truthful report, whereas the other workers truthfully report their gender, that is \( \{p^*, q^*, s^*\} = \{1,0,0\} \) and \( \{p^*, q^*, s^*\} = \{0,0,1\} \);
- **c.** if \( \frac{c_x}{\beta_x \Delta u} \geq \frac{1}{2} \), there exists an equilibrium in which all workers truthfully report their gender, that is, \( \{p^*, q^*, s^*\} = \{1,0,1\} \).

**Proof of Proposition 5:** Consider a putative equilibrium in which \( \{p^*, q^*, s^*\} = \{0,0,0\} \).

Given \( q^* = s^* = 0 \), from (23), \( U_1(x,x) - U_1(x,y) < 0 \) (i.e., \( p^* = 0 \)) if \( \frac{c_y}{\beta_y \Delta u} < \frac{1}{3} - \frac{2}{3} \beta_u \beta_y \). Given \( p^* = s^* = 0 \), from (24), \( U_2(y,x) - U_2(y,y) < 0 \) (i.e., \( q^* = 0 \)) if \( \frac{c_y}{\beta_y \Delta u} < \frac{1}{3} - \frac{2}{3} \beta_u \beta_y \). Given \( p^* = q^* = 0 \), from (25), \( U_3(x,x) - U_3(x,y) < 0 \) (i.e., \( s^* = 0 \)) if \( \frac{c_x}{\beta_x \Delta u} < \frac{1}{3} - \frac{2}{3} \beta_u \beta_y \), which finally proves point (a) in the Proposition.

Consider now a putative equilibrium in which \( \{p^*, q^*, s^*\} = \{1,0,0\} \). Given \( q^* = s^* = 0 \), from (23), \( U_1(x,x) - U_1(x,y) \geq 0 \) (i.e., \( p^* = 1 \)) if \( \frac{c_x}{\beta_x \Delta u} \geq \frac{1}{3} - \frac{2}{3} \beta_u \beta_y \). Given \( p^* = 1 \) and \( s^* = 0 \),

\footnote{For the sake of crispness, we have here deviated from the rule according to which the indifference between a truthful and a non-truthful report is broken in favor of the former.}
from (24), \( U_2(y,x) - U_2(y,y) < 0 \) (i.e., \( q^* = 0 \)). Given \( p^* = 1 \) and \( q^* = 0 \), from (25), \( U_3(x,x) - U_3(x,y) < 0 \) (i.e., \( s^* = 0 \)) if \( \frac{c_x}{\beta_y \Delta u} < \frac{1}{2} \). Similar steps show that \( \{p^*, q^*, s^*\} = \{0,0,1\} \) is also an equilibrium for \( \frac{c_x}{\beta_y \Delta u} \in \left[ \frac{1}{3} - \frac{2}{3} \frac{\beta_u}{\beta_y}, \frac{1}{2} \right] \), which finally proves point \((b)\) in the Proposition.

Consider now a putative equilibrium in which \( \{p^*, q^*, s^*\} = \{1,0,1\} \). Given \( q^* = 0 \) and \( s^* = 1 \), from (23), \( U_1(x,x) - U_1(x,y) \geq 0 \) (i.e., \( p^* = 1 \)) if \( \frac{c_x}{\beta_y \Delta u} \geq \frac{1}{2} \). Given \( p^* = 1 \) and \( s^* = 1 \), from (24), \( U_2(y,x) - U_2(y,y) < 0 \) (i.e., \( q^* = 0 \)). Given \( p^* = 1 \) and \( q^* = 0 \), from (25), \( U_3(x,x) - U_3(x,y) < 0 \) (i.e., \( s^* = 1 \)) if \( \frac{c_x}{\beta_y \Delta u} > \frac{1}{2} \), which finally proves point \((c)\) in the Proposition.

Proposition 5 shows that the workers’ equilibrium behavior described in Propositions 1 and 2 continues to hold when we introduce a type of \( P \) with a preference for under-represented genders. As mentioned above, the main difference with the equilibrium behavior described in Section 3 lies in the fact that workers whose real gender is less likely to be discriminated \( \text{per se} \) (i.e., \( g = y \)) may not have a dominant strategy. Like Propositions 1 and 2, workers whose real gender is more likely to be discriminated \( \text{per se} \) (i.e., \( g = x \)) react by hiding their gender.

Although each of the equilibria described in Proposition 5 have a corresponding equilibrium in Propositions 1 and 2, the probability that a worker whose reported gender is \( x \) is hired differs between the two settings. In particular, from (4) and (22), it is easily seen that, all else equal (i.e., keeping constant workers’ equilibrium choices), the probability that \( P \) hires a worker whose reported gender is \( x \) is higher when \( \beta_u \) is strictly positive rather than when it is
equal to zero. In this sense, the presence of a firm with preference for under-represented genders leads to a lower degree of discrimination on the labor market, without fundamentally affecting workers’ choices.

---

49 We are aware that the existence of a given equilibrium depends on $\beta_u$ (see Proposition 6). Yet, this statement holds true for values of the parameters such that the same equilibrium behavior arises in both settings (when $\beta_u > 0$ and $\beta_u = 0$). For instance, this is the case when $\frac{c_x}{\beta_y} \in \left[\frac{1}{3}, \frac{1}{2}\right]$ (see Propositions 1-2 and 5).

50 When $\beta_y > \beta_x = 0$, introducing $\beta_u > 0$ can also lead to a lower probability that $A_1$ and $A_3$ hide their gender, thereby further decreasing the degree of discrimination observed on the labor market. For instance, if $\frac{c_x}{\beta_y} \in \left[\frac{1}{3} - \frac{2}{3} \frac{\beta_u}{\beta_y}, \frac{1}{2}\right]$, and for sufficiently high values of $c_y$, from (23)-(24)-(25), there exists an equilibrium in which $\{p^*, q^*, s^*\} = \left\{\frac{2\beta_y - \beta_u + \frac{3c_x}{\beta_u}}{\beta_y + 2\beta_u}, 0, \frac{2\beta_y - \beta_u + \frac{3c_x}{\beta_u}}{\beta_y + 2\beta_u}\right\}$, where $\{p^*, s^*\}$ is increasing in $\beta_u$. When $\frac{c_x}{\beta_y} \in \left[\frac{1}{3}, \frac{1}{2}\right]$, we can compare this equilibrium to that reported in Proposition 2, that is, $\{p^*, q^*, s^*\} = \left\{2 \left[\frac{3c_x}{\beta_y} - 1\right], 0, 2 \left[\frac{3c_x}{\beta_y} - 1\right]\right\}$, where the two equilibria are identical for $\beta_u = 0$. 

75
Appendix J. Post-experimental survey

Table 8. Post-Experimental Survey

Panel A: Female and male respondents’ beliefs and willingness to pay

<table>
<thead>
<tr>
<th></th>
<th>Male performance</th>
<th>Female performance</th>
<th>Male workers hired</th>
<th>Female workers hired</th>
<th>Willingness to pay to be male</th>
<th>Willingness to pay to be female</th>
<th>Under-represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT (30)</td>
<td>15.57</td>
<td>15.07</td>
<td>56.87%</td>
<td>43.13%</td>
<td>4.97</td>
<td>3.47</td>
<td>46.67%</td>
</tr>
<tr>
<td>MNIT (31)</td>
<td>-</td>
<td>-</td>
<td>55.29%</td>
<td>44.71%</td>
<td>4.55</td>
<td>2.79</td>
<td>61.29%</td>
</tr>
<tr>
<td>EIT (30)</td>
<td>9.93</td>
<td>11.60</td>
<td>43.13%</td>
<td>56.87%</td>
<td>3.03</td>
<td>4.20</td>
<td>36.67%</td>
</tr>
</tbody>
</table>

Panel B: Male respondents’ beliefs and willingness to pay

<table>
<thead>
<tr>
<th></th>
<th>Male performance</th>
<th>Female performance</th>
<th>Male workers hired</th>
<th>Female workers hired</th>
<th>Willingness to pay to be male</th>
<th>Willingness to pay to be female</th>
<th>Under-represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT (15)</td>
<td>15.20</td>
<td>14.73</td>
<td>53.87%</td>
<td>46.13%</td>
<td>4.00</td>
<td>3.67</td>
<td>33.33%</td>
</tr>
<tr>
<td>MNIT (15)</td>
<td>-</td>
<td>-</td>
<td>52.67%</td>
<td>47.33%</td>
<td>4.13</td>
<td>2.87</td>
<td>53.33%</td>
</tr>
<tr>
<td>EIT (15)</td>
<td>9.87</td>
<td>11.07</td>
<td>42.27%</td>
<td>57.73%</td>
<td>2.60</td>
<td>4.20</td>
<td>33.33%</td>
</tr>
</tbody>
</table>

Panel C: Female respondents’ beliefs and willingness to pay

<table>
<thead>
<tr>
<th></th>
<th>Male performance</th>
<th>Female performance</th>
<th>Male workers hired</th>
<th>Female workers hired</th>
<th>Willingness to pay to be male</th>
<th>Willingness to pay to be female</th>
<th>Under-represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT (15)</td>
<td>15.93</td>
<td>15.40</td>
<td>59.87%</td>
<td>40.13%</td>
<td>5.93</td>
<td>3.27</td>
<td>60.00%</td>
</tr>
<tr>
<td>MNIT (16)</td>
<td>-</td>
<td>-</td>
<td>57.75%</td>
<td>42.25%</td>
<td>4.94</td>
<td>2.72</td>
<td>68.75%</td>
</tr>
<tr>
<td>EIT (15)</td>
<td>10.00</td>
<td>12.13</td>
<td>44.00%</td>
<td>56.00%</td>
<td>3.47</td>
<td>4.20</td>
<td>40.00%</td>
</tr>
</tbody>
</table>

Note: The table reports the respondents’ beliefs about (i) the performance of male and female participants in the laboratory experiment in the mathematical and emotion-recognition tasks (columns 1 and 2), (ii) the percentage of male and female workers hired in the three different treatments (columns 3 and 4), (iii) the percentages of firms with a preference for under-represented genders in the pool of workers (column 7). The table also reports the respondents’ willingness to pay to be perceived as either a male or a female in our laboratory setting in the three different treatments (columns 5 and 6). The participants in this post-experimental survey are students at the University of Exeter, different from those who participated in the laboratory experiment. We report the number of observations in brackets next to the treatment acronym (e.g., in Panel B, MIT (15) indicates that the number of observations is 15).