Does a stereotype benefit women in the labor market: An experiment on perseverance

Simone Haeckl and Melis Kartal*

December 9, 2021

Abstract

We design a novel real-effort experiment to investigate gender differences and stereotypes regarding perseverance and how these affect employment decisions. We find that women are more persistent than men and that the subjects anticipate this difference. While it pays off, in expectation, to hire a female over a male candidate in our experimental employment setting, employers are not more likely to hire a female candidate. Thus, even in a setting where female candidates are statistically better workers and employers hold a positive stereotype about women, employers do not hire women more often than men. This finding contrasts with studies showing that men do benefit from positive stereotypes associated with them, and suggests that stereotypes might be more beneficial for men than for women.

Keywords: non-cognitive skills, perseverance, conscientiousness, gender and stereotypes, labor market experiment

JEL classification: J16, J71, C91,

*Haeckl: University of Stavanger, Department of Economics; simone.haeckl-schermer@uis.no;
Kartal: Vienna University of Economics and Business (WU Vienna), Department of Economics; melis.kartal@wu.ac.at.

We thank the WULABS, Vienna University of Economics and Business and Vienna Center for Experimental Economics (VCEE) for allowing us to run the experiments in their laboratory. We gratefully acknowledge financial support by WULABS. We appreciate helpful comments from Ben Greiner, Gergely Hajdu, Mari Rege, Rupert Sausgruber, and Robert Stueber.

1. Introduction

Many surveys and studies suggest that females are (perceived to be) more diligent and persistent than males.[1] Differences in non-cognitive skills, such as perseverance, are important to labor market outcomes, as labor and personnel economics has established that non-cognitive skills are important determinants of labor market success (Almlund et al., 2011).[2] In fact, according to Bowles et al. (2001), employers in low-skill labor markets value traits such as perseverance more than cognitive ability (see also Heckman et al., 2006). Still, gender differences in non-cognitive skills have received much less attention in the literature than differences in cognitive and mathematical skills, which seem to put men at an advantage (Guiso et al., 2008).

We argue that ignoring gender differences in non-cognitive skills might lead to an underestimation of gender gaps. This underestimation might be even more pronounced if there are stereotypes about skills in addition to actual skill differences.[3] For example, in the context of cognitive and mathematical skills, various studies have documented negative stereotypes about women and their effect on women's educational choices and labor market success.[4] In contrast, we have limited knowledge of how women benefit from positive stereotypes concerning non-cognitive skills.[5] If such stereotypes exist and are not taken into account, this could have important implications for the magnitude of the (unexplained) gender gaps. For example, low-skill jobs are characterized by a significant gender wage gap (Card et al., 2015; Kochhar, 2020). This is puzzling because non-cognitive skills such as perseverance, in which women might have a (perceived) advantage over men, are particularly important for these types of jobs, as mentioned above.

To gain insight into the impact of perseverance and stereotypes concerning perseverance on gender differences in employment decisions, we designed a laboratory experiment. We use this experiment to (i) obtain a measure of subjects' perseverance; (ii) test whether there are gender stereotypes concerning perseverance; and (iii) investigate whether gender

[1]See Balart and Oosterveen (2019), Christensen and Knezek (2014), Duckworth and Seligman (2006), Flinn et al. (2020), and Jõgi and Kikas (2016)

[2]Non-cognitive skills may be broadly defined as personality traits or "patterns of thought, feelings, and behavior" that are weakly correlated with measures of intelligence (Brunello and Schlotter, 2011).

[3]Based on the definition of stereotypes provided by Hilton and Von Hippel (1996), "stereotypes are beliefs about the characteristics, attributes, and behaviors of members of certain groups". Therefore, a stereotype might be a valid generalization or an exaggeration of existing differences (see also Bordalo et al., 2016).

[4]See for example Bohren et al. (2019b), Bordalo et al. (2019), Carlana (2019), Carli et al. (2016), Coffman (2014), Dutt et al. (2016), Günther et al. (2010), Hirnstein et al. (2014), Reilly (2012), Reuben et al. (2014), and Steele et al. (2002).

[5]There is, however, some evidence for positive discrimination toward women in female-dominated labor markets from correspondence studies (see for example Ahmed et al., 2021; Booth and Leigh, 2010).

differences in perseverance and stereotypes regarding these differences affect employment decisions.

Perseverance can be defined as continued effort and persistence in doing something despite difficulties or delayed success (throughout this paper, we use 'persistence' and 'perseverance' interchangeably). To measure perseverance, the design combines a *real-effort task* in which subjects try to complete tasks under time pressure with a paid *leisure* alternative as an *outside option*. Subjects are rewarded for their effort only if they manage to complete the task on time, whereas the leisure alternative pays them a small flat rate per second. As completing the task requires a diligent effort, and the subjects have an attractive outside option (i.e., paid leisure), we use the time subjects spend working on the task as our measure of perseverance.

In addition to actual gender differences in perseverance, we investigate whether there are perceived gender differences, i.e., whether there is a stereotype. We elicit subjects' beliefs about the perseverance of men and women in our task.[6] Based on the survey-based literature discussed above, we expect that women are more likely to persist in our task than men, and that subjects perceive females as the more persistent gender.

Building on these expectations, we investigate whether gender-based stereotypes affect employment decisions using a design based on Reuben et al. (2014). After the real-effort task is over, subjects in the role of an employer decide to "hire" one of two "candidates". In each round, two subjects are randomly selected to be the candidates. The remaining subjects have the role of employers in this round. The incentives for the employer are designed in such a way that, in order to maximize earnings, the employer must select the candidate who they believe is more persistent. The hiring procedure is repeated until every subject has assumed the role of candidate once. We use the strategy method, and one randomly chosen hiring decision determines the employer's earnings.

We vary the information provided to employers between treatments to investigate whether information can alter the impact of a stereotype on hiring decisions. In the *control* treatment, employers only see the candidates when they stand up and announce their randomly assigned candidate numbers, but receive no additional information. In the *info* treatment, employers make two hiring decisions in each round. The initial decision in the *info* treatment is similar to the decision in the *control* treatment: employers only see the current round's candidates and have no other information about them. This decision is used to identify learning effects, i.e., whether information from previous rounds affects sub-

[6]Beliefs are elicited only at the end of the experiment, to reduce the salience of the focus of the experiment on gender.

sequent decisions. After the initial decision, the employers receive information about the previous work effort of the candidates and can revise their hiring decision. We use treatment differences to investigate whether information has an impact on hiring decisions.

In our design, we find that women persist more and are expected to do so by a majority of subjects based on the belief elicitation task results.[7] In fact, we also find a positive stereotype about women's skill in the task although we did not expect it ex-ante, and it is not validated by the data. However, employers are not more likely to hire a woman rather than a man. In other words, the stereotypes that we find do not translate into increased hiring of women. This lack of a boost in female hiring despite two positive stereotypes is surprising, since there is ample evidence that men benefit from stereotypes in employment settings even in the absence of actual gender differences. (Barron et al., 2020; Bohren et al., 2019a; Coffman et al., forthcoming; Reuben et al., 2014).

Regarding the treatment variation, we find no difference in employment decisions between treatments. One explanation for this finding is that the information signal provided to employers in the *info* treatment (i.e., candidates' previous work effort) underpredicts the gender difference in perseverance we observe in later work phases. In particular, the information signal provided to employers does not reflect the fact that women reduce their work effort much less than men over time. However, we find that even in those sessions in which the signals are highly informative regarding the gender difference, their impact on employment choices is quite limited.

2. Theoretical framework and contribution to the literature

We now discuss in more formal terms why perseverance may matter in the labor market. Consider a stylized setting in which the productivity of individual i, denoted by p_i, is given by $p_i = e_i + a_i + \epsilon_i$, where $e_i \in [0, \bar{e}]$ is the effort choice by i, $a_i \in (0, \bar{a}]$ is an ability parameter, and ϵ_i is a stochastic luck parameter.[8] Let the cost of effort equal $k_i e_i^2$, where $k_i > 0$. Given the definition of perseverance, our interpretation is that more perseverance is associated with a smaller value of k, which reduces the cost of effort, all else equal. The productivity p_i determines $\pi(p_i) \in \{0, 1\}$, i.e., whether or not i reaches a milestone on time (for example, timely completion of a job task or project). In particular, $\pi(p_i) = 1$ and i obtains a reward of $w > 0$ if and only if p_i is higher than a certain threshold $\bar{p}$. Thus, a

[7] Note that we elicit beliefs only in the control group, which does not receive feedback about other participants' perseverance. See Figure 1 for details.

[8] For simplicity, we assume additivity of a and e. This assumption is, however, not necessary and our discussion could be generalized to any well-behaved function $f(a, e)$ strictly increasing in a and e.

higher level of perseverance (i.e., a lower value of k_i) increases the optimal effort choice e_i and makes it more likely to reach the milestone, all else equal.

The key variables of this framework are a and k, representing the cognitive ability and perseverance, respectively. These variables are related to gender gaps in labor market outcomes, stereotypes, and employer choices.[9] As discussed above, it is well-documented that stereotypes can be relevant to the gender gap in labor market outcomes. In the context of our stylized model, the stereotypes concern the employer's belief about the values of a and k for each employee. For example, there may be a stereotype that, on average, a is greater for men than women. If such a stereotype is largely an exaggeration, then this may (among other things) lead to inefficient employment decisions as employers choose men over women more often than they should.

Our study focuses instead on k, the perseverance parameter, which may "advantage" women (Balart and Oosterveen, 2019; Christensen and Knezek, 2014; Duckworth and Seligman, 2006; Flinn et al., 2020; Jõgi and Kikas, 2016).[10] Based on the suggestive evidence for gender differences in perseverance, we expect women to exhibit more perseverance than men: that is, they have a lower k. In addition, we hypothesize that there is a stereotype concerning this difference: that is, the subjects believe that women are more persistent than men. Provided that we find evidence for a stereotype placing women at an advantage, we also expect to find that women are hired more frequently than men for tasks requiring perseverance.

While the stereotype that women are more persistent should give them an advantage in hiring decisions, it is important to note that, in theory, it may have a negative effect on their wages. For example, the simple framework we outlined above implies that if the employer knows that for women, k is lower, they can be offered a lower wage than men for the same effort.[11] We elaborate more on this point and future directions for research in the concluding section.

[9] While this paper focuses on the behavior of the employers, there is a growing literature showing that stereotypes impair the confidence and performance of the "disadvantaged" group (Bordalo et al., 2019; Coffman, 2014; Günther et al., 2010; Steele et al., 2002).

[10] A very recent strand of the literature shows that women are less likely than men to continue "competing" after failure. In contrast to this paper, those studies concern high-level competition, such as a math Olympiads (Buser and Yuan, 2019), a speed-cube solving competition (Fang et al., 2021), and applications to a highly selective scientific conference (Pereda et al., 2020). Hence, they measure perseverance in highly competitive settings where cognitive skills are focal. Women tend to be averse to the feedback they receive from failing in such settings (see for example Bertrand, 2011). Instead, our experimental task is non-competitive and poses no threat to the ego. This is in order for our design to be able to cleanly identify the role of persistent and diligent effort.

[11] Assume that the employer can observe effort and offers the minimum acceptable wage, which is the minimum w that allows the employee to reach the reservation utility r at the effort level e_i stipulated by the employer. Then, for fixed r, w is lower for lower k.

Because women lag behind men in labor market outcomes and wages, they are the "disadvantaged" group in the literature on gender gaps and gender stereotypes (see for example Bertrand, 2011; Buser et al., 2014; Buser and Yuan, 2019; Croson and Gneezy, 2009; Falk and Hermle, 2018, and the references therein). Put differently, the literature so far has focused on cognitive skills and traits which result in men being perceived as the "advantaged" group (e.g., the group with the higher a). We focus instead on perseverance, an important non-cognitive skill, which may advantage women. This advantage should be particularly pronounced in low-skill jobs and/or when labor demand is short. However, we are not aware of any study documenting a female advantage in the labor market due to a positive stereotype about their skills. While we document the existence of a positive stereotype about women, this does not translate to increased chances of being hired, which is in contrast to previous evidence showing that men benefit from positive stereotypes even when there is no actual gender difference.

Our paper is also related to the growing literature in economics and psychology that has systematically documented the importance of non-cognitive skills and personality traits as predictors of health, education, productivity, labor market success, and financial outcomes (Almlund et al., 2011; Borghans et al., 2008; Bowles et al., 2001; Brunello and Schlotter, 2011; Cubel et al., 2016; Heckman and Rubinstein, 2001; Kautz et al., 2014; Lindqvist and Vestman, 2011; Moffitt et al., 2011; Roberts et al., 2014, and the references therein).[12] We contribute to this literature by investigating gender differences and stereotypes regarding non-cognitive skills and their impact on employment decisions.

Lastly, our paper also relates to the experimental economic literature on perseverance. There are two laboratory studies on perseverance that we are aware of: Buechel et al. (2018) and Gerhards and Gravert (2020). Both measure perseverance in a lab experiment but, unlike our paper, focus on peer effects in perseverance. They find evidence of a peer effect: observing a persistent person makes the observer try harder. Perseverance is also related to a personality trait called "grit" in the recent psychology literature.[13] Grit is defined as "perseverance and passion for long-term goals" (Duckworth et al., 2007) and has been shown to be positively correlated with economic success (Duckworth et al., 2007; Duckworth and Quinn, 2009). In a meta-study, Credé et al. (2017) emphasize that this correlation is largely due to the perseverance facet of grit. Alan et al. (2019) design and

[12]See Bowles and Gintis (1976) and Edwards (1976) for an earlier discussion of the importance of non-cognitive skills.

[13]See Alaoui and Fons-Rosen (2021) for a discussion of how perseverance is related to grit and conscientiousness.

evaluate a randomized education intervention aimed at fostering grit among school children, showing that grit is malleable.

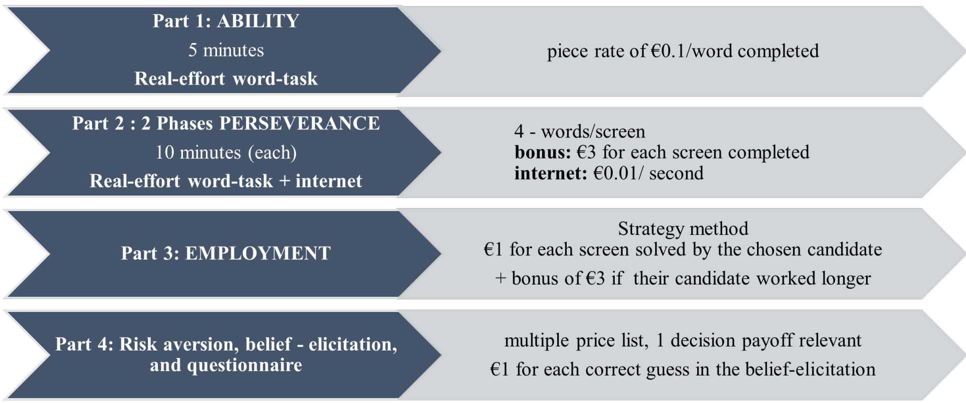
3. Experimental design

Using a laboratory experiment, we investigate (perceived) gender differences in perseverance and their influence on employment choices. We begin by describing the experimental schedule and its components in detail. Then, we discuss the experimental treatments. Lastly, we outline the experimental protocol.

3.1. Experimental schedule

The experiment consists of four parts, as outlined in Figure 1. We will now describe each part in detail in chronological order.

Figure 1: Experimental schedule



Part 1: Ability

Part 1 is a practice phase in which the subjects work on individual anagrams. That is, for five minutes, the subjects reorder strings of six letters to build an English word, and receive a piece rate of ten cents per correctly solved anagram. We ensure that the subjects understand the instructions by including a control screen with an anagram that everyone has to complete before the start of Part 1. Subjects see only one anagram per screen, there is no paid leisure option, and there is no time limit per screen, except for the total time limit of five minutes. In the regression analysis, we use the output from this part as a proxy to control for each subject's ability to solve anagrams.

Part 2: Perseverance

Part 1 is followed by Part 2, the perseverance part. This part consists of two identical phases ("perseverance phases") in which the subjects work on the "perseverance task" described below. Each phase lasts for ten minutes and there is a one-minute break in between the two phases.

After the subjects read the instructions and before they begin the first perseverance phase, we ask them to guess their performance in that phase. We use their answers to control for differences in beliefs, which may also explain why some subjects work harder than others apart from differences in perseverance.

To measure perseverance, we have developed a design combining a real-effort task that requires persistence with a paid leisure option. In particular, the subjects are presented with a screen displaying a set of four anagrams, that is, four strings of six letters, each of which can be reordered to form an English word. In contrast to the ability part, there is a time limit for each screen. Subjects have 60 seconds to solve all four anagrams on a screen and receive a bonus of €3 if they correctly complete the entire screen. After the 60 seconds expire or after the subject clicks on a button indicating they are finished with or want to skip their screen, the procedure starts again with a new screen displaying a new set of anagrams. This procedure continues until the time limit of the perseverance phase is up.

The subjects can take a break from the task and spend time on surfing the Internet ("leisure"). For each second spent on the Internet, they receive €0.01. The subjects know that they can go back and forth between the real-effort task and leisure as frequently as they wish.[14] The instructions clearly explain these features of the experiment in order to rule out the possible influence of the "social desirability of work." During the instructions, the subjects also go through a brief training period to ensure that they understand how to switch back and forth between the task screen and the Internet browser (see the instructions in Appendix D).

As it is difficult (but possible) to finish a screen within the limited time frame, it is necessary for most subjects to persist in the task, rather than surf the Internet, in order to increase their chances of solving a screen successfully.[15] Therefore, we use the time subjects spent working on the real-effort task as a proxy for perseverance. Using the time spent on trying to solve anagrams as a measure of perseverance is a cleaner measure than output, as it is not confounded with differences in ability, and is well established in the psychology

[14] The subjects were aware that they are not allowed to use the Internet to solve the anagrams. Additionally, we blocked pages that can be used to solve anagrams.

[15] To find a good level of difficulty, we ran a pretest in the lab of the Vienna Center for Experimental Economics.

literature. The earliest such measure that we are aware of dates back to Eisenberger and Leonard (1980). Variations of anagram tasks have been used to measure persistence by, among others, Sandelands et al. (1988), Quinn et al. (1996), Brandon et al. (2003), Nussbaum and Steele (2007), and Lucas et al. (2015). However, unlike in our experiment, many studies in psychology have deceived subjects and used impossible anagrams (that is, the subjects were unaware that the anagrams they worked on had no solution). Our tasks are solvable but still require perseverance for successful completion. In addition, our measure is incentivized, and we employ a paid leisure option to prevent a possible experimenter demand effect and the related social desirability of work discussed above.[16]

Based on the suggestive evidence for gender differences in perseverance discussed in the Introduction, we form our first two hypotheses:

Hypothesis 1: *Women spend more time working in the perseverance phases than men.*

Hypothesis 2: *Subjects believe that women work more in the perseverance phases than men, i.e., there is a stereotype advantaging women.*

Part 3: Employment decisions

The perseverance part is followed by the employment part. The design of this part is based on Reuben et al. (2014). There are several identical rounds, and in each round, subjects who are in the role of an employer make individual hiring decisions. At the beginning of a round, two subjects are randomly selected (without replacement) to be in the role of the candidates.[17] These two subjects stand up and declare the "candidate number" assigned to them. The other subjects are in the role of an employer and decide which of the two candidates to hire for that round. Employers make their hiring decisions privately on their computers by ticking the box next to the number of their chosen candidate. The procedure continues until every subject has been in the role of a candidate once. We apply this more organic procedure to present the candidates instead of providing employers with male and female avatars or names on the computer screen since we want to understate the gender aspect of the study. For the same reason, our procedure employs both same-gender and

[16] Since we focus on the time spent in the perseverance task and the level of difficulty is similar across the task screens, our method differs from Gerhards and Gravert (2020), who use an anagram task to proxy perseverance with the number of anagrams subjects skip and with the number of times subjects switch from hard and high-reward anagrams to easy but low-reward anagrams. Our design allows us to measure the skipping behavior of subjects. We show the robustness of our results with respect to an alternative perseverance measure based on skipping behavior in Appendix B. Similar to our design, the measure of perseverance in Buechel et al. (2018) is the time participants spend on a "depleting" real-effort task. Our design differs from theirs in that we use anagram tasks and allow subjects to switch back and forth between anagrams and leisure.

[17] The total number of rounds in this part depends on the number of subjects (n) in a session and equals $\frac{n}{2}$.

mixed-gender pairs in the employment part and asks subjects about gender-specific beliefs only at the end of the experiment. The order of the same- and mixed-gender pairs is constant across all sessions.

At the end of this part, one round is randomly chosen for payment. The payoff to the employer depends on the task effort of the selected candidate in the second perseverance phase as follows: the employer receives €3 if the selected candidate spent more time working on the task in the second perseverance phase than the other candidate and €1 for each screen the selected candidate completed within the time limit. The subjects are not informed of the amount of payoff they have earned until the end of the experiment. We base this payoff not only on work time but also on the number of completed screens, since a payoff measure that relies on both effort and output is more realistic.

Note that since employment decisions take place after the perseverance part, and the subjects were not aware of the employment part while working on the real-effort task, the social preferences of candidates (especially of female candidates) cannot have affected the provision of the effort. Also, subjects do not receive a payoff in their role as candidates in the employment part in order to prevent social preferences from influencing employers' hiring decisions because we are interested in learning which of the two candidates an employer believes will make them better off.

Provided that we find evidence for Hypothesis 2 (i.e., there is evidence for a stereotype placing women at an advantage), we expect to find evidence for the following hypothesis:

Hypothesis 3: *Women are hired more frequently than men in the employment part.*

Part 4: Beliefs

In Part 4, we elicit the subjects' beliefs about whether men or women worked more in the second perseverance phase in Part 2 (stereotype about perseverance). We also ask the subjects to guess whether men or women solved more screens in Part 1 of the experiment (stereotype about ability). The elicitation of these beliefs is incentive-compatible, as the subjects earn €1 for each correct guess. We interpret the elicited beliefs of the subjects concerning the perseverance task as an explicit measure of gender stereotypes about perseverance, whereas employment decisions provide us with an implicit measure. We place the elicitation of beliefs on the possible gender differences in the last part of the experiment, in order to reduce the salience of gender in the employment decisions in Part 3. We also elicit the risk preferences of the subjects using a lottery-based multiple-price list with constant probabilities, as suggested by, among others Drichoutis and Lusk (2016). Lastly, subjects answer a short questionnaire on their demographics as well as on their perception of the real-effort task, and receive feedback on their earnings.

3.2. Treatments

We investigate and compare employment decisions in two treatments: *control* and *info*. These treatments differ only in Part 3 and employ a between-subject design. Part 3 of the *control* treatment proceeds exactly as described above: the employers see the two candidates as they announce their candidate numbers, and must make the employment decision without any other information. In the *info* treatment, employers make two decisions in each round. The initial decision of a round is analogous to the decision in the *control*. After the initial decision, employers are informed about the work time of the candidates in the first perseverance phase. While the work time in the first perseverance phase is not relevant per se to the payoff to the employers, the two phases have the same incentive structure and therefore, the work time in the first perseverance phase may be a valuable signal regarding the work effort of the candidates in the second phase. After receiving this information, the employers have the chance to revise their employment decision without any cost. We ask the participants in *info* to make two decisions, because the initial decision allows us to see if the information about specific pairs in previous rounds affects the initial decisions, i.e., whether there is learning.

In the first version of the experiment, only the second hiring decision in the info treatment was relevant to the payoff (Sessions 1 and 3). We changed this feature in later treatment sessions (sessions 6 and 8) and randomly selected one of the two hiring decisions in the round to determine the payoff to the employer in Part 3.[18] The subjects were informed about this procedure before making their hiring decisions. In Appendix B, we show that this change does not have an effect on the employment decisions. Therefore, we pool the data from both sets of sessions in the main part of this paper.

3.3. Experimental protocol

We conducted our experiment at the laboratory of the Vienna University of Economics and Business (WULABS) using z-tree (Fischbacher, 2007). The subjects were recruited from a student sample using ORSEE (Greiner, 2015). There were 194 students, participating in 8 sessions. On average, each session lasted 90 minutes, and the subjects earned €16.11.

[18] Such a change is not relevant for the control treatment, since the subjects made only one hiring decision.

We aimed for a gender-balanced sample and invited the same number of male and female students in each session. Females made up 53% of the subjects, and the share of female subjects per session ranged from 50%–60%.

4. Results

We first discuss the gender difference in our behavioral measure of perseverance and test Hypothesis 1. We then, to test Hypothesis 2, investigate the perceived gender differences based on the elicited beliefs of subjects. Next, we analyze the employment decisions to test whether they are consistent with the elicited beliefs and the actual performance by gender (Hypothesis 3). Lastly, we discuss the effect of information regarding past performance on employment decisions. We use one-sided nonparametric tests for directed hypotheses and two-sided tests otherwise.

4.1. Gender differences in perseverance

As discussed in the previous section, we use the time subjects spend working on the real-effort task as our behavioral measure of perseverance. On average, the respective work time on the task was 376 seconds and 302 seconds in the first and the second perseverance phases (out of 600 seconds each). The difference between the two phases is statistically significant ($p < 0.001$, Wilcoxon sign-rank test) and points to a *wear-out effect*, which is not surprising, as the task is demanding and requires persistence.

Table 1: Women are more persistent than men

Work time in the perseverance Task				
	Phase 1	Phase 2	Change	Total
Women (N = 102)	387.26 (427.12)	333.62 (366.48)	-53.64 ($p < 0.001$)	720.88 (800.21)
Men (N = 92)	363.69 (382.89)	267.40 (244.60)	-96.30 ($p < 0.001$)	631.09 (579.03)
Difference	23.57 ($p = 0.152$)	66.22 ($p = 0.014$)	42.65 ($p = 0.014$)	89.79 ($p = 0.004$)

Notes: The mean work time in seconds (the median in parentheses). The p-values are based on Wilcoxon–Mann–Whitney tests, gender comparisons are based on one-sided tests as we have directional hypotheses.

Table 1 shows the mean and median work times by gender. In both perseverance phases, women spend more time working on the task than men. This gender difference is significant in the second perseverance phase, as well as in the combined data of the two phases (i.e., the total work time in the task). In the first perseverance phase, women spend, on average, 387 seconds working, while men work for 364 seconds ($p = 0.152$, one-sided Wilcoxon–Mann–Whitney test). In the second phase, the difference in work time by gender increases to 66 seconds ($p = 0.014$, one-sided Wilcoxon–Mann–Whitney test). Combining the data of both phases, the total work time difference is 90 seconds ($p = 0.004$, one-sided Wilcoxon–Mann–Whitney test).

As Table 1 shows, men decrease their work time by nearly twice as much as women do. Across the two phases, men reduce their work effort by 96 seconds (26%) in the second phase, whereas women reduce it by 54 seconds (14%). The gender difference in the reduction of effort is statistically significant ($p = 0.014$, one-sided Wilcoxon–Mann–Whitney test).

Figure 2: There are two different types of workers

Notes: The left panel shows the kernel density of the work time in the first perseverance phase for females (black lines) and males (gray lines). The right panel shows the density of the work time in the second phase. All subjects are included in the graph.

In the second perseverance phase, two extreme types of workers emerged: those who work very little, and those who work throughout the whole of the phase. Figure 2 shows the distribution of the time spent working on the task, separately by gender, for the two perseverance phases. In the second phase (right panel in Figure 2), 38% of the subjects either work less than one minute (loafers) or spend less than one minute surfing the Internet (worker bees). These two types are distributed very unequally by gender. While 63% of the loafers are male, 62% of the worker bees are female ($p = 0.061$, Fisher's exact test testing for equality of the share of loafers and worker bees by gender). As expected, work times in the two perseverance phases are highly correlated, i.e., people who work more in phase 1 are also more likely to work more in phase 2 (pairwise correlation coefficient 0.74, $p < 0.001$).

Apart from differences in perseverance, work-time differences could be driven by differences in the ability to solve anagrams and differences in confidence in this ability, both of which influence the expected return to effort in the perseverance task. Regarding ability, there is no evidence of a gender difference. Recall that our measure of ability is the number of anagrams solved in Part 1, where a subject receives a piece rate for each correctly solved anagram, and there is no leisure option. On average, women solve 10.18 anagrams, and men solve 9.35. This difference is insignificant ($p = 0.451$, Wilcoxon–Mann–Whitney test). We do *not* use the number of completed screens in the perseverance task in Part 2 to measure ability because, as we will explain below, work time in Part 2 affects task performance, and it differs between men and women.

To study the effect of the work time on task performance in Part 2, we run a Poisson regression in which the number of completed screens in the perseverance phases is the dependent variable, and the time spent working on the task is the explanatory variable.[19]

Table 2: Number of completed screens in Part 2 increases in work time

	Number of completed screens in the perseverance phases (Part 2)		
	(1) Phase 1	(2) Phase 2	(3) Total
Work time in phase 1	1.005***		
	(0.001)		
Work time in phase 2		1.006***	
		(0.001)	
Total work time			1.003***
			(0.000)
Ability	1.089***	1.095***	1.088***
	(0.018)	(0.021)	(0.016)
Constant	0.024***	0.017***	0.038***
	(0.010)	(0.008)	(0.013)
Observations	194	194	194

Notes: We use a Poisson regression to account for the structure of the data. Coefficients show incident rate ratios. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2 shows that according to this regression, the number of completed screens increases in the work time. For example, considering the regression with the total work time

[19] In Table A2 in Appendix A, we show that the results do not qualitatively change when using an OLS regression instead.

as the independent variable, if the total work time increases by one second, the incidence rate ratio increases by a factor of 1.003 keeping all else equal (see Column 3). Put differently, an increase in the total work time by 333.33 seconds increases the expected number of completed screens by one screen on average. As women work more in the perseverance phase, we test if they also solve more screens. Across the two perseverance phases of Part 2, women complete on average 1.16 screens while men complete 0.80 (p = 0.075, one-sided Wilcoxon–Mann–Whitney test, see also Table A1 in Appendix A). Looking at the two perseverance phases separately, men and women complete 0.40 and 0.54 screens respectively in the first phase (p = 0.306, one-sided Wilcoxon–Mann–Whitney test). In the second phase, the gender difference increases, as women reduce their work effort less than men did: men and women complete 0.40 and 0.62 screens, respectively (p = 0.043, one-sided Wilcoxon–Mann–Whitney test). The small gender difference in output is entirely driven by gender differences in work time. Adding a gender dummy to Table 2 does not evince a significant effect (see Table A3).

To summarize, women solve more screens than men by spending more time on the real-effort task, but as a consequence, they also give up earnings from paid leisure. The two countervailing effects cancel out, and men and women eventually do not differ in earnings. The total earnings in Part 2 for women and men are, respectively, €8.26 and €8.10 (p = 0.498, Wilcoxon–Mann–Whitney test). Looking at the two phases of Part 2 separately, the earnings for women and men are, respectively, €3.75 and €3.57 in the first phase (p = 0.950, Wilcoxon–Mann–Whitney test), and €4.52 and €4.53 in the second phase (p = 0.405, Wilcoxon–Mann–Whitney test).

While we argue that the gender difference in work time is driven by a difference in perseverance, two other potential explanations are gender differences in confidence and risk aversion. Our measure of confidence is the number of screens that a subject ex-ante expects to complete in the two perseverance phases.

Men are more confident than women and believe on average that they will solve 3.26 screens, whereas women believe that they will solve 2.77. This difference is statistically significant, based on a Wilcoxon–Mann–Whitney test (p = 0.041). As subjects receive a bonus only if they are able to finish a screen on time in the perseverance task but earn a flat rate when surfing the Internet, a gender difference in risk aversion could also explain why one gender works more than the other. However, we do not find evidence for this explanation, as women are more risk averse than men but spend more time on the ("riskier") perseverance task. We use the number of choices of a risky option in a price list as our proxy for risk aversion. Women choose the risky option on average four times while men choose the risky option five times ($p = 0.022$, Wilcoxon-Mann-Whitney test).

We run a Tobit regression to identify the determinants of work time. Table 3 presents the results of this regression. We find that the average work time in the second perseverance phase as well as the total work time are significantly higher for women than men, controlling for ability, confidence, and risk aversion. The average work time also increases in ability (recall that ability does not differ by gender). Neither confidence nor risk-aversion influences work time.

Table 3: Work time is higher for women, controlling for ability, confidence, and risk aversion

	Work time in the perseverance phases (Part 2)		
	(1) Phase 1	(2) Phase 2	(3) Total
Female	24.850	71.432**	162.507**
	(24.843)	(31.600)	(59.897)
Ability	5.357**	10.062***	16.052***
	(2.521)	(3.133)	(6.678)
Confidence	3.279*	-1.271	15.663
	(1.932)	(1.383)	(11.843)
Risk aversion	3.463	7.891	12.620
	(5.023)	(5.838)	(12.433)
Constant	287.224***	139.690***	336.663***
	(35.066)	(39.808)	(88.464)
Observations	194	194	194

Notes: Tobit regressions with the time subjects spent working as the dependent variable, which is by design censored at 0 and 600. Ability is proxied by the number of anagrams solved in part 1. Confidence shows subjects' ex-ante belief about how many screens they will be able to complete in the perseverance phase. Risk aversion is proxied by the number of risky decisions taken in a multiple price list. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Combining our findings in Part 2, we conclude that the evidence supports Hypothesis 1. Not only do women spend more time working on the task than men overall across the two phases and in the second phase separately, but they also reduce their work effort less than men do in the second phase, consistent with the definition of perseverance.

The gender difference in perseverance is also anticipated by the subjects. We only consider the control group because the beliefs are elicited after the employment decisions and by this stage, the participants in the info treatment have received information about the candidates' provision of effort. In the control group, 59% of the participants believe that men spent more time surfing the Internet, which is marginally significantly different from

0.5 ($N = 94$, $p = 0.049$ one-sided test of proportions).[20] Thus, we conclude that there is support for Hypothesis 2: subjects believe that women work more in the perseverance phases than men. In addition, 62% of our subjects believe that women solved more anagrams in Part 1 of the experiment ($N = 94$, $p = 0.023$ two-sided test of proportions). Hence, we find a positive stereotype about women's ability to solve anagrams, although this stereotype is not validated by the data.

4.2. Employment decisions

In Part 3, an employer earns €3 if they choose the candidate who spent more time working and €1 for each screen the chosen candidate completed (both measured in the second perseverance phase of Part 2). Since there is a direct monetary incentive for selecting the candidate who spent more time working, and the number of completed screens increases in work time, as shown in Table 2, employers should simply choose the candidate who they believe worked longer. As discussed above, a majority of the participants believe that women work more (and that women are better at the task, although this is not supported by the data). Hence, we expect to find that women are also hired more. Before we investigate hiring decisions, we check whether the gender differences in work time and screens solved, discussed in the previous section, translate to a higher payoff from hiring a female candidate. Excluding the 21 same-gender pairs, we find that the female candidate worked longer than the male candidate in 54% of the pairs (based on work times in the second perseverance phase). Looking at the payoff each candidate in a mixed-gender pair would have generated for their employer if they had been chosen, women generate on average a payoff of €2.13 while men generate an average payoff of €1.89. However, this difference is not statistically significant ($p = 0.253$, $N = 152$, one-sided Wilcoxon–Mann–Whitney test).[21]

As there are many same-gender pairs, for the sake of reducing the salience of the gender aspect of the study, we lose a significant amount of data. However, since the pairs of candidates were matched randomly, we can randomly re-match all subjects into (same- and mixed-gender) pairs and extend the data analysis. We repeat the re-sampling procedure 1000 times to get a distribution of estimates of gender differences. For example, the results of our re-sampling method show that the proportion of mixed-gender pairs in which the female candidate works longer than the male candidate is 60% on average (across all replications) and larger than 50% in 932 of our 1000 randomly re-matched samples. This difference

[20] In the Conclusion, we address the potential concern that social desirability might affect stated beliefs more than employment choices .

[21] In all, we have 76 mixed-gender pairs and therefore compare 152 paired subjects.

results in a gender difference in the payoff from hires. In mixed-gender pairs, the average payoff generated by female candidates is €2.41, and the average payoff generated by male candidates is €1.61 across all randomly re-matched samples. Hiring the female candidate pays off, on average, in 60% of the employment decisions. In 935 out of 1000 samples the share of females generating the higher payoff is larger than 50%. We conclude that in our experiment, it pays to hire a female candidate, on average.

In Table 4, we examine the employers' hiring decisions in Part 3 of our experiment and address Hypothesis 3. The female candidate is hired in 53% of the mixed-gender pairs in the *control* treatment, which is significantly lower than the predicted 60% of decisions in which choosing the female candidate pays off.

Table 4: There is no treatment effect on employment decisions

	Share of female candidates chosen
Info (initial decision)	-0.014
	(0.041)
Info (after information)	0.002
	(0.087)
Constant	0.533***
	(0.034)
Observations	1722
Number of employers	194
p-values of Wald tests for share = 0.60%	
Control treatment	0.049
Info Treatment (initial)	<0.001
Info Treatment (after info)	0.416

Notes: Results based on a linear probability model taking into account the panel structure of the data using employer random effects. The control treatment is our baseline. The same-gender pairs of candidates are excluded from the analysis. Numbers inside parentheses are standard errors clustered by session.[22] * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Concerning treatment effects, the fraction of females hired is not affected by the *info* treatment, and the impact of information signals that advantage female candidates is muted, as we discuss in more detail below in Section 4.3. Hence, we conclude that there is no support for Hypothesis 3: although the expected payoff generated by women is higher than that for men and there is a stereotype that women work more than men, the fraction of female candidates preferred over male candidates does not match the expected payoff.[23]

[22]The results do not change significantly if we use robust standard errors instead.

[23]In fact, there is also a positive stereotype about women's ability in the task, as mentioned before (but it is not validated by the data).

4.3. Treatment effects and learning

Table 4 shows that the share of women hired does not vary significantly between treatments (Table A4 in Appendix A replicates Table 4 using a probit model). This statement holds independently of whether we consider the initial decision of a round (i.e., before receiving information about the candidates' work time in the first perseverance phase) or the second decision of the round (*control* (53%) versus initial decision of a round in *info* (52%): $p = 0.787$, Wald-test; *control* (53%) versus second decision in *info* (53%): $p = 0.821$, Wald-test). While this result may seem surprising at first, it is mainly driven by the fact that the work time in the first perseverance phase underrepresents the gender difference in the second perseverance phase as already shown in Figure 2. Looking at the performance signals employers receive in the *info* treatment, we see that in only 51% of the cases does the employer receive a signal that the female candidate worked more. In addition, a large majority (85%) of the employers in the *info* treatment hire the candidate who worked more in the first perseverance phase. Thus, there is no treatment effect on the aggregate level.

To investigate whether learning takes place conditional on receiving signals that predominantly advantage women (or men) in the mixed pairs of previous rounds, we make use of the between-session variation.[24] In Figure 3, we split the sample into two types of sessions: four sessions which predominantly generated information signals that the female candidate worked more in the first perseverance phase (female-signal) and four sessions which generated either roughly balanced signals or predominantly signals that the male candidate worked more (male signal). It is important to note that not only do the female-signal sessions provide employers with a higher number of signals that the female candidate worked more (relative to the number of signals that the male candidate worked more in male-signal sessions), but they also provide predominantly female signals in the early periods which might be particularly important for learning (see Figure A1 in Appendix A for the cumulative number of female signals per period).

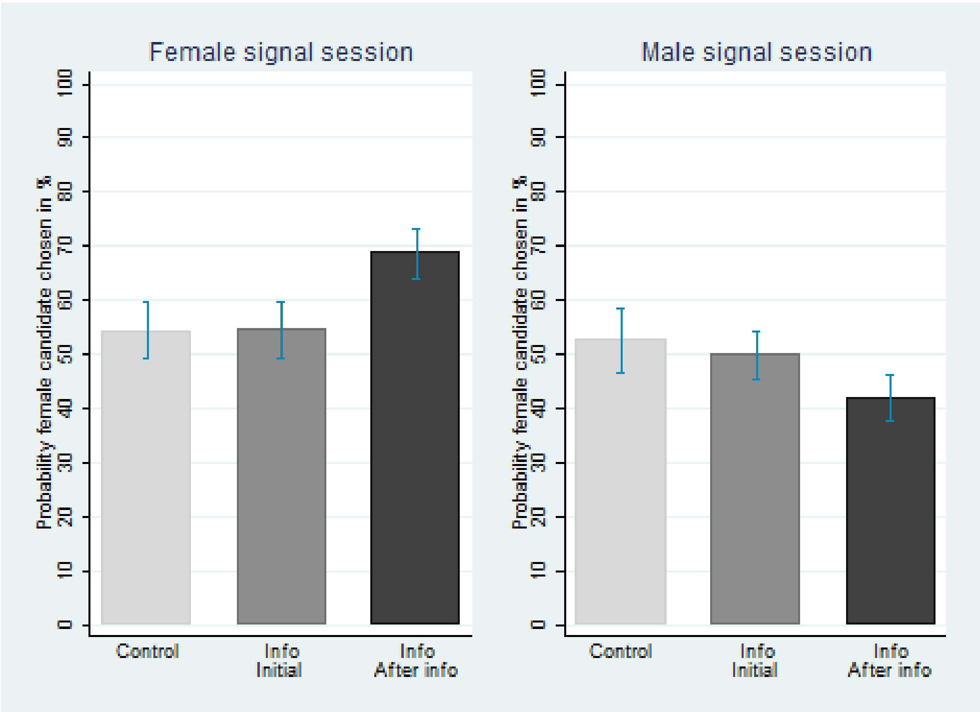
Figure 3 presents the share of women hired in the *control* treatment according to our classification of sessions (male- or female-signal) as a benchmark. As expected, the share of female candidates hired by employers does not differ between female- and male-signal sessions in the *control* treatment since no information is provided in this treatment.[25] In the second decision of a round in the *info* treatment, employers often follow the information they received regarding the candidates' previous work effort and are significantly more likely

[24]See Table A5 for an overview of work times per session.

[25]Comparing the light gray bars in the left and right panels of Figure 3, we see that the female candidate is hired in 54% of the cases in female-signal sessions, and in 53% of the cases in the male-signal sessions ($p = 0.719$, Wald test, see also Tables A6 and A7 in Appendix A).

to employ the female candidate in the female signal sessions: comparing the black bars in the left and right panels of Figure 3, we see the share of women hired decreases from 69% in the female signal sessions to 42% in the male signal sessions ($p = 0.003$, Wald test). These numbers seem to suggest that there is an asymmetry in updating from signals, which favors female candidates. However, this asymmetry is due to the fact that the share of female signals in female-signal sessions is larger than the share of male signals in the male-signal sessions, from early rounds on, as indicated before.

Figure 3: There is little learning in the info treatment, controlling for signals.



Notes: The shares and confidence intervals are based on margins from a random-effect probit regression taking into account the panel structure of the data by clustering standard errors on the employer level (see table A7). In the left panel only the four sessions (two in the *control* and two in the *info* treatment) in which employers received predominantly a signal that the female candidate worked more are included. The number of decisions is therefore 394 in the *control* and 428 in the *info* treatment. In the right panel only the four sessions (two in the *control* and two in the *info* treatment) in which employers received predominantly the signal that the male candidate worked more are included. The number of decisions is therefore 480 in the *control* and 572 in the *info* treatment. The light gray bars represent the *control* treatment and the dark gray and the black bars the *info* treatment.

With regard to learning from previous rounds' information signals and its impact on the initial hiring decision of a round in the *info* treatment, we do not find a significant difference between the shares of employed women in female- and male-signal sessions. Comparing the dark gray bars in Figure 3, the proportion of women hired in mixed-gender pairs is 55% and 50% respectively in female- and male-signal sessions (p = 0.587, Wald test). We conclude that employers do not update from the information signals they received in earlier rounds. This is surprising, since female-signal sessions provide employers with predominantly female signals from early rounds on.

To summarize, we find evidence that women persist more than men in a task that requires diligent effort even if there is an attractive leisure option. Furthermore, we find that this gender difference is anticipated by employers, which, surprisingly, does not translate into a higher chance of choosing a female candidate over a male candidate. In addition, the employers updating using information signals seems limited to information about the the current round's candidates: that is, employers do not seem to update beliefs based on previous rounds' performance signals.[26]

5. Conclusion

Using a novel real-effort experiment, we investigate gender differences and stereotypes in perseverance and the relevance of these factors to employment decisions. We find that women are more persistent than men and that subjects anticipate this difference. This study finds two ways in which women are more persistent than men. Firstly, women work more than men in the two perseverance phases, and secondly, they reduce their effort less than men across phases. We argue that this cannot be explained by gender differences in the ability to solve anagrams or in confidence.

We define a stereotype about perseverance to be the belief that women have spent more time in the perseverance phase. A potential concern with testing for the existence of a stereotype using our belief-elicitation method is that gender is more salient during belief elicitation than during the employment decision, and participants might, therefore, try to answer in the "socially desirable" way. However, it is not clear why it should be socially desirable to state that women work longer than men and the belief elicitation is monetarily incentivized. In addition, the elicited beliefs are in line with the employees' actual behavior and simulation data.

One explanation for why women work harder may be that women make sub-optimal choices between effort and leisure in our experiment, and hence, they "overwork". However, there is no statistically significant gender difference in earnings in our experiment (despite the fact that leisure brings a small monetary return). In addition, according to our theo-

[26] In Appendix C, we also conduct an exploratory analysis investigating how employment decisions and beliefs differ by the employer's gender.

retical framework, being diligent reduces the cost of effort. As a result, if women are more diligent than men, utility maximization is associated with women's choices of higher levels of effort than those chosen by men, in line with what we observed in the experiment.

In relation to this, we conjecture that gender differences in perseverance, which we argue are associated with differences in the cost of effort, can help explain gender differences in "reservation wages " observed in the literature (Caliendo et al., 2017). For example, the simple framework we outlined in Section 2 implies that, assuming that the employer offers a reservation wage and knows that women have a lower cost of effort, women are offered a lower wage than men for the same effort, all else equal. In line with this notion, Litman et al. (2020) find that women are, on average, willing to work for a lower wage in an online marketplace (MTurk). They show that these differences cannot be explained by differences in productivity, job selection, or discrimination by the employers, but may be driven by lower reservations wages. In turn, according to our conjecture above and our experimental findings, women have lower reservation wages than men because they have more perseverance and a lower cost of effort. As Litman et al. (2020) exclude employer discrimination by design, the resulting gender wage gap is probably a lower bound and could be even more pronounced if employers were aware that women are willing to exert the same effort at a lower wage than men.

More generally, our study is related to studies of various gender gaps in the labor market, as the magnitude of the unexplained gender gaps may be underestimated if there is a skill advantage or a positive stereotype about the perseverance of females that is not taken into account. While employing a female candidate has a higher expected return in our experiment, employers are not more likely to hire female candidates unless they learned about the individual performances of the candidates in a previous working period. Thus, even in a setting where female employees have a justifiable advantage in employment decisions, which is also anticipated (there is even an additional positive stereotype about women that is not validated), employers might not favor women over men.

References

Ahmed, Ali, Mark Granberg, and Shantanu Khanna (Jan. 2021). "Gender Discrimination in Hiring: An Experimental Reexamination of the Swedish Case". *PLOS ONE* 16.1, pp. 1–15.

Alan, Sule, Teodora Boneva, and Seda Ertac (2019). "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit". *The Quarterly Journal of Economics* 134.3, pp. 1121–1162.

Alaoui, Larbi and Christian Fons-Rosen (2021). "Know when to fold'em: The flip side of grit". *European Economic Review* 136, p. 103736.

Almlund, Mathilde, Angela L. Duckworth, James J. Heckman, and Tim Kautz (2011). "Personality Psychology and Economics". *Handbook of the Economics of Education* 4, pp. 1–181.

Balart, Pau and Matthijs Oosterveen (2019). "Females Show More Sustained Performance During Test-Taking than Males". *Nature Communications* 10.1, p. 3798.

Barron, Kai, Ruth Ditlmann, Stefan Gehrig, and Sebastian Schweighofer-Kodritsch (Aug. 2020). *Explicit and Implicit Belief-Based Gender Discrimination: A Hiring Experiment.* Discussion Papers, Research Unit: Economics of Change SP II 2020-306. WZB Berlin Social Science Center.

Bertrand, Marianne (2011). "New Perspectives on Gender". *Handbook of Labor Economics.* Ed. by David Card and Orley Ashenfelter. Vol. 4. Elsevier, pp. 1543–1590.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope (2019a). *Inaccurate Statistical Discrimination: An Identification Problem.* Working Paper 25935. National Bureau of Economic Research.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg (2019b). "The Dynamics of Discrimination: Theory and Evidence". *American Economic Review* 109.10, pp. 3395–3436.

Booth, Alison and Andrew Leigh (2010). "Do Employers Discriminate by Gender? A Field Experiment in Female-Dominated Occupations". *Economics Letters* 107.2, pp. 236–238.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2016). "Stereotypes". *The Quarterly Journal of Economics* 131.4, pp. 1753–1794.

– (2019). "Beliefs about Gender". *American Economic Review* 109.3, pp. 739–773.

Borghans, Lex, Angela Duckworth, James Heckman, and Bas Weel (Feb. 2008). "The Economics and Psychology of Personal Traits". *The Journal of Human Resources* 43.

Bowles, Samuel and Herbert Gintis (1976). *Schooling in Capitalist America.* New York: Basic Books.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne (2001). "The Determinants of Earnings: A Behavioral Approach". *Journal of Economic Literature* 39, pp. 1137–1176.

Brandon, Thomas H., Thaddeus A. Herzog, Laura M. Juliano, Jennifer E. Irvin, Amy B. Lazev, and Vani Nath Simmons (2003). "Pretreatment task persistence predicts smoking cessation outcome." *Journal of Abnormal Psychology* 112.3, p. 448.

Brunello, Giorgio and Martin Schlotter (2011). "Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education & Training Systems". *IZA Discussion Paper* No. 5743.

Buechel, Berno, Lydia Mechtenberg, and Julia Petersen (2018). "If I Can Do It, So Can You! Peer Effects On Perseverance". *Journal of Economic Behavior & Organization* 155, pp. 301–314.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek (2014). "Gender, competitiveness, and career choices". *The Quarterly Journal of Economics* 129.3, pp. 1409–1447.

Buser, Thomas and Huaiping Yuan (2019). "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad". *American Economic Journal: Applied Economics* 11.3, pp. 225–52.

Caliendo, Marco, Lee Wang-Sheng, and Robert Mahlstedt (2017). "The Gender Wage Gap and the Role of Reservation Wages: New Evidence for Unemployed Workers". *Journal of Economic Behavior & Organization* 136, pp. 161–173.

Card, David, Ana Rute Cardoso, and Patrick Kline (2015). "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women". *The Quarterly Journal of Economics* 131.2, pp. 633–686.

Carlana, Michela (2019). "Implicit Stereotypes: Evidence from Teachers' Gender Bias". *The Quarterly Journal of Economics* 134.3, pp. 1163–1224.

Carli, Linda L., Laila Alawa, YoonAh Lee, Bei Zhao, and Elaine Kim (2016). "Stereotypes About Gender and Science". *Psychology of Women Quarterly* 40.2, pp. 244–260.

Christensen, Rhonda and Gerald Knezek (2014). "Comparative Measures of Grit, Tenacity and Perseverance". *International Journal of Learning, Teaching and Educational Research* 8, pp. 16–30.

Coffman, Katherine B. (2014). "Evidence on Self-Stereotyping and the Contribution of Ideas". *The Quarterly Journal of Economics* 129.4, pp. 1625–1660.

Coffman, Katherine B., Christine L. Exley, and Muriel Niederle (forthcoming). "The Role of Beliefs in Driving Gender Discrimination". *Management Science*.

Credé, Marcus, Michael C. Tynan, and Peter D. Harms (2017). “Much Ado About Grit: A Meta-Analytic Synthesis of the Grit Literature”. *Journal of Personality and Social Psychology* 113.3, pp. 492–511.

Croson, Rachel and Uri Gneezy (2009). “Gender Differences in Preferences”. *Journal of Economic Literature* 47.2, pp. 448–74.

Cubel, Maria, Ana Nuevo-Chiquero, Santiago Sanchez-Pages, and Marian Vidal-Fernandez (2016). “Do Personality Traits Affect Productivity? Evidence from the Laboratory”. *The Economic Journal* 126.592, pp. 654–681.

Drichoutis, Andreas C. and Jayson L. Lusk (2016). “What Can Multiple Price Lists Really Tell Us About Risk Preferences?” *Journal of Risk and Uncertainty* 53.2-3, pp. 89–106.

Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly (2007). “Grit: Perseverance and Passion for Long-Term Goals”. *Journal of Personality and Social Psychology* 92, pp. 1087–1101.

Duckworth, Angela L. and Patrick D. Quinn (2009). “Development and Validation of the Short Grit Scale (Grit-S)”. *Journal of Personality Assessment* 91, pp. 166–174.

Duckworth, Angela L. and Martin E. P. Seligman (2006). “Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores”. *Journal of Educational Psychology* 98.1, pp. 198–208.

Dutt, Kuheli, Danielle L. Pfaff, Ariel F. Bernstein, Joseph S. Dillard, and Caryn J. Block (2016). “Gender Differences in Recommendation Letters for Postdoctoral Fellowships in Geoscience”. *Nature Geoscience* 9.11, pp. 805–808.

Edwards, Richard C (1976). “Individual Traits and Organizational Incentives: What Makes a ”Good” Worker?” *Journal of Human Resources*, pp. 51–68.

Eisenberger, Robert and Janet M. Leonard (1980). “Effects of Conceptual Task Difficulty on Generalized Persistence”. *The American Journal of Psychology* 93.2, pp. 285–298.

Falk, Armin and Johannes Hermle (2018). “Relationship of Gender Differences in Preferences to Economic Development and Gender Equality”. *Science and Society* 362.6412.

Fang, Chao, Ernest Zhang, and Junfu Zhang (2021). “Do Women Give Up Competing More Easily? Evidence from Speedcubers”. *Economics Letters* 205, p. 109943.

Fischbacher, Urs (2007). “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments”. *Experimental Economics* 10.2, pp. 171–178.

Flinn, Christopher, Petra Todd, and Weilong Zhang (2020). *Personality Traits, Job Search and the Gender Wage Gap*. Tech. rep. Human Capital and Economic Opportunity Working Group.

Gerhards, Leonie and Christina Gravert (2020). “Because of You I Did Not Give Up – Peer Effects in Perseverance”. *Journal of Economic Psychology* 81, p. 102316.

Greiner, Ben (2015). “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE”. *Journal of the Economic Science Association* 1.1, pp. 114–125.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales (2008). “Culture, Gender, and Math”. *Science* 320.5880, pp. 1164–1165.

Günther, Christina, Neslihan A. Ekinci, Christiane Schwieren, and Martin Strobel (2010). “Women Can’t Jump?—An Experiment on Competitive Attitudes and Stereotype Threat”. *Journal of Economic Behavior & Organization* 75.3, pp. 395–401.

Heckman, James J. and Yona Rubinstein (2001). “The Importance of Noncognitive Skills: Lessons from the GED Testing Program”. *American Economic Review* 91.2, pp. 145–149.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006). “The Effects of Cognitive and Noncognitive Abilities on Labor Market - Outcomes and Social Behavior”. *Journal of Labor Economics* 24.3, pp. 411–482.

Hilton, James L. and William Von Hippel (1996). “Stereotypes”. *Annual Review of Psychology* 47.1, pp. 237–271.

Hirnstein, Marco, Lisa Coloma Andrews, and Markus Hausmann (2014). “Gender-Stereotyping and Cognitive Sex Differences in Mixed- and Same-Sex Groups”. *Archives of Sexual Behavior* 43.8, pp. 1663–1673.

Jõgi, Anna-Liisa and Eve Kikas (2016). “Calculation and Word Problem-Solving Skills in Primary Grades – Impact of Cognitive Abilities and Longitudinal Interrelations with Task-Persistent Behaviour”. *British Journal of Educational Psychology* 86.2, pp. 165–181.

Kautz, Tim, James J Heckman, Ron Diris, Bas ter Weel, and Lex Borghans (2014). *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success.* Working Paper 20749. National Bureau of Economic Research.

Kochhar, Rakesh (2020). *Women Make Gains in the Workplace Amid a Rising Demand for Skilled Workers.* Tech. rep. Pew Research Center.

Lindqvist, Erik and Roine Vestman (2011). “The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment”. *American Economic Journal: Applied Economics* 3.1, pp. 101–28.

Litman, Leib, Jonathan Robinson, Zohn Rosen, Cheskie Rosenzweig, Joshua Waxman, and Lisa M. Bates (2020). “The Persistence of Pay Inequality: The Gender Pay Gap in an Anonymous Online Labor Market”. *PLOS ONE* 15.2.

Lucas, Gale M., Jonathan Gratch, Lin Cheng, and Stacy Marsella (2015). "When the Going Gets Tough: Grit Predicts Costly Perseverance". *Journal of Research in Personality* 59, pp. 15–22.

Moffitt, Terrie E., Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W. Roberts, Stephen Ross, Malcolm R. Sears, W. Murray Thomson, and Avshalom Caspi (2011). "A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety". *Proceedings of the National Academy of Sciences* 108.7, pp. 2693–2698.

Nussbaum, David A. and Claude M. Steele (2007). "Situational Disengagement and Persistence in the Face of Adversity". *Journal of Experimental Social Psychology* 43.1, pp. 127–134.

Pereda, Paula, Liz Matsunaga, Maria D. M. Diaz, Bruna P. Borges, Jesus Mena-Chalco, Fabiana Rocha, Renata Narita, and Clara Brenck (Sept. 2020). *Are Women Less Persistent? Evidence from Submissions to a Nationwide Meeting of Economics.* Working Papers, Department of Economics 2020$_1$9. University of Sao Paulo (FEA-USP).

Quinn, Edward P., Thomas H. Brandon, and Amy L. Copeland (1996). "Is Task Persistence Related to Smoking and Substance Abuse? The Application of Learned Industriousness Theory to Addictive Behaviors". *Experimental and Clinical Psychopharmacology* 4.2, pp. 186–190.

Reilly, David (2012). "Gender, Culture, and Sex-Typed Cognitive Abilities". *PLOS ONE* 7.7.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2014). "How Stereotypes Impair Women's Careers in Science". *Proceedings of the National Academy of Sciences* 111, pp. 4403–4408.

Roberts, Brent W., Carl Lejuez, Robert F. Krueger, Jessica M. Richards, and Patrick L. Hill (2014). "What is Conscientiousness and How Can it Be Assessed?" *Developmental Psychology* 50, pp. 1315–1330.

Sandelands, Lloyd E., Joel J. Brockner, and Mary A. Glynn (1988). "If at 1st You Dont Succeed, Try, Try Again - Effects of Persistence Performance Contingencies, Ego Involvement, and Self-Esteem on Task Persistence". *Journal of Applied Psychology* 73.2, pp. 208–216.

Steele, Claude M., Steven J. Spencer, and Joshua Aronson (2002). "Contending with Group Image: The Psychology of Stereotype and Social Identity Threat". Advances in Experimental Social Psychology 34, pp. 379–440.

A. Additional Tables

Table A1: Screens completed in the second perseverance phase

	Women	Men	Difference	p-value
Phase 1	0.54	0.40	-0.13	0.306
Phase 2	0.62	0.40	-0.22	0.043
Total	1.16	0.80	-0.35	0.075

Notes: Mean number of completed screens in the perseverance part. p-values are based on one-sided Wilcoxon–Mann–Whitney test.

Table A2: Number of completed screens in Part 2 increases in work time

	Number of completed screens in the Perseverance phases (Part 2)		
	(1) Phase 1	(2) Phase 2	(3) Total
Work time in phase 1	0.002***		
	(0.000)		
Work time in phase 2		0.002***	
		(0.000)	
Total work time			0.002***
			(0.000)
Ability	0.054***	0.071***	0.123***
	(0.015)	(0.018)	(0.027)
Constant	-0.648***	-0.766***	-1.532***
	(0.190)	(0.180)	(0.320)
Observations	194	194	194

Notes: OLS regression with the aggregate number of completed screens as the dependent variable. Coefficients show incident rate ratios. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A5: Work times in perseverance per session

	Women		Men	
	Phase1 1	Phase 2	Phase 1	Phase 2
Session1	437.56 (554.21)	429.89 (592.09)	281.74 (281.98)	169.84 (133.17)
Session2	342.34 (416.40)	250.66 (197.60)	407.71 (459.27)	286.02 (293.47)
Session3	431.23 (439.47)	298.84 (339.20)	311.43 (369.28)	209.22 (143.46)
Session4	412.40 (437.23)	386.53 (424.67)	326.06 (365.92)	237.58 (196.48)
Session5	388.08 (425.95)	335.78 (381.53)	398.19 (483.47)	329.31 (401.95)
Session6	331.14 (381.97)	261.22 (289.74)	408.28 (481.47)	303.18 (279.02)
Session7	399.55 (398.41)	337.98 (228.73)	343.96 (389.10)	314.80 (294.30)
Session8	342.56 (351.76)	248.61 (410.55)	421.67 (421.06)	288.70 (268.44)

Notes: Mean work time in seconds, medians in parentheses.

Table A3: Number of completed screens in Part 2 increases in work time

Number of completed screens in the perseverance phases (Part 2)			
	(1) Phase 1	(2) Phase 2	(3) Total
Work time in phase 1	1.005***		
	(0.001)		
Work time in phase 2		1.006***	
		(0.001)	
Total work time			1.003***
			(0.000)
Gender	0.932	0.951	0.940
	(0.197)	(0.238)	(0.172)
Ability	1.088***	1.094***	1.087***
	(0.017)	(0.021)	(0.015)
Constant	0.025***	0.017***	0.040***
	(0.010)	(0.009)	(0.009)
Observations	194	194	194

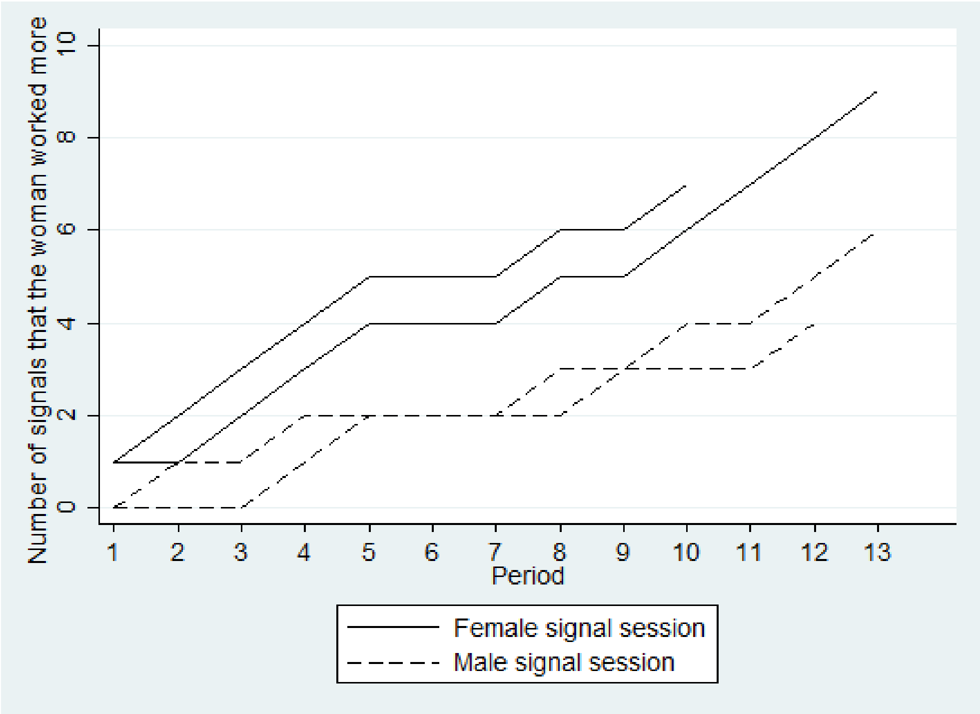
Notes: We use a Poisson regression to take into account the structure of the data. Coefficients show incident rate ratios. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A4: Treatment differences in employment decisions (Probit)

		Share female candidate chosen	
Info (initial decision)		-0.027	
		(0.068)	
Info (after information)		0.038	
		(0.050)	
Constant		0.084	
		(0.053)	
Observations		194	
	Predicted Probabilities	95% Confidence Interval	
Control treatment	0.534	0.491	0.575
Info Treatment (initial)	0.520	0.483	0.557
Info Treatment (after info)	0.542	0.502	0.582

Notes: Probit regression taking into account the panel structure of the data using employer random effects. The control treatment is the baseline. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Figure A1: Cumulative number of female signals for female and male signal sessions



Notes: The solid lines represent the cumulative number of signals that the female candidate worked more in the female signal sessions, the dashed line represents the cumulative number for male signal sessions. Only sessions in the *info* treatment are included, as this is the only treatment in which the subjects got information about the candidates' previous performance.

Table A6: Treatment differences in employment decisions dependent on session composition

	Share female candidate chosen
Info (initial decision)	-0.025
	(0.037)
Info (after information)	-0.081
	(0.019)
Female signal session	0.015
	(0.039)
Info (initial decision) X Female signal session	0.026
	(0.053)
Info (after information)X Female signal session	0.224***
	(0.031)
Constant	0.525***
	(0.020)
Observations	1722
Number of employers	194

Notes: Results based on a linear probability model taking into account the panel structure of the data using employer random effects. The control treatment in male sessions is the baseline. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A7: Treatment differences in employment decisions dependent on session composition (Probit)

	Share female candidate chosen		
Info (initial decision)		-0.063	
		(0.103)	
Info (after information)		-0.221*	
		(0.117)	
Female signal sessions		0.040	
		(0.111)	
Info (initial decision) X Female signal sessions		0.064	
		(0.153)	
Info (after information)X Female Signal sessions		0.598***	
		(0.177)	
Constant		0.066	
		(0.076)	
Observations		1722	
Number of employers		194	
	Predicted Probabilities female signal	95% Confidence Interval	
Control treatment	0.545	0.493	0.595
Info Treatment (initial)	0.545	0.492	0.597
Info Treatment (after info)	0.688	0.641	0.732
	Predicted Probabilities male signal	95% Confidence Interval	
Control treatment	0.526	0.465	0.586
Info Treatment (initial)	0.500	0.455	0.545
Info Treatment (after info)	0.419	0.377	0.461

Notes: Random-effect probit regression taking into account the panel structure of the data using employer random effects. The control treatment in male sessions is the baseline. Numbers in parentheses are standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A8: Beliefs about gender differences in work time by gender

	Share believing women worked more
Total work time	-0.000574
	(0.000402)
Male employer	0.617**
	(0.270)
Constant	0.332
	(0.345)
Observations	94

Notes: Probit regression with a dummy indicating that the subject believed that women worked more as the dependent variable. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

B. Robustness checks

How does our measure of perseverance relate to the measure proposed by Gerhards and Gravert (2020)?

As discussed in Section 3, Gerhards and Gravert (2020) proxy perseverance by the number of skipped anagrams (the more anagrams a subject skips, the less persistent he or she is). We are also able to observe skipping behavior if we use the data from part 1 (ability), because, in part 1, subjects received a piece rate for each correctly solved anagram, but could skip as many anagrams as they liked. We find that this proxy for perseverance is significantly correlated with the time subjects spent surfing the Internet (leisure). Leisure is the complement to the time they spent working in the two perseverance phases, which we use as a proxy for perseverance (the Pearson correlation coefficient between tasks skipped and leisure in the first perseverance phase are 0.23, $p < 0.001$ and 0.26 in the second perseverance phase, $p < 0.001$).

Table B1: Determinants of the time spent surfing the Internet (leisure)

	Leisure time in perseverance phases	
	(1)	(2)
	Phase 1	Phase 2
Skips in Part 1	9.056***	10.439***
	(2.282)	(2.700)
Belief success perseverance phase	-2.978	1.459
	(2.331)	(1.673)
Male	17.851	63.394**
	(22.961)	(30.932)
Constant	135.862***	157.898***
	(28.291)	(34.260)
Observations	194	194

Notes: Tobit regression with the time spent surfing the Internet in the two perseverance phases as dependent variable. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B1 shows Tobit regressions with leisure time in the part 2 (perseverance phases) as the dependent variable. Column (1) shows leisure in phase 1, column (2) leisure in phase 2 as the dependent variables. In line with the correlation coefficients discussed above, we find that leisure in both phases is significantly correlated with the skipping behavior in part 1. Interestingly, we also find this, even controlling for skipping behavior. Men spent significantly more time in leisure than women in phase 2 (+ 63 seconds, $p =$ 0.040, see column 2). This indicates that our measure covers a slightly different aspect of perseverance. As discussed in Section 3, our measure captures the willingness to continue working hard even after repeated setbacks.

Robustness to changes in the payment for the employment decisions

After the first four sessions ("old sessions"), we changed the incentives provided to the employers in the *info* treatment. While in the old sessions, only the second decision, i.e., the decision they made after receiving information about the candidates' performance in the first perseverance phase, was used to determine the payoff, in the new sessions, one of the two decisions was picked at random to determine these payoffs. To test if this change affected the employers' decisions, we compare their switching behavior between the no info and the info choice across sessions. Table B2 shows the share of employers changing their decisions. If the lack of incentives for the initial decision in the old sessions caused subjects to think less about which candidate to choose, they should be revising their decisions more frequently than the subjects in the new sessions. We do not find evidence for such a change in behavior. In the old sessions, employers revised 32% of their decisions, in the new sessions they revised 30% ($p = 0.585$, Fisher's exact test).

Table B2: Share of employers changing decisions

	old sessions	**new sessions**	**Total**
Share not changing the decision	68.07	69.53	68.85
Share changing the decision	31.93	30.47	31.15

Fisher's exact test = 0.585

In Table B3, we use random-effect probit regressions to investigate if the probability that the employer picked the candidate who worked more in their initial decision varies between the old and the new sessions (column 1). We cannot reject the hypothesis that the shares of employers picking the candidate who indeed worked more varies between sessions. In column (2), we look again at the share of employers changing their decisions and do not find a significant difference between old and new sessions. We, therefore, conclude that we can pool the data from both sets of sessions.

Table B3: Decisions are not affected by the change in incentives

	(1) Chosen candidate worked more (initial decision)	(2) Changed decision after information
New session	0.929	1.009
	(0.071)	(0.104)
Constant	1.030	0.596***
	(0.0054)	(0.046)
Observations	100	100

Notes: Probit regression taking into account the panel structure of the data by using employer random effects. Only subjects in the *Info* treatment are included. Numbers in parentheses are robust standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C. Gender differences in beliefs and employment decisions

In Section 4, we found that men are more likely to hire the female candidate than are women. This gender difference is driven by differences in beliefs.

While 71% of men in the control group believed that women on average worked more in the perseverance phases, only 47% of the women did ($p = 0.018$, test of proportions). Moreover, 71% of the men believed that women are better at solving anagrams than men are, while only 53% of the women did so ($p = 0.072$, test of proportions). This difference in beliefs is not driven by introspection, as the men are significantly more likely to believe that women work more than men even if we control for their own work time (see Table A8).

The gender difference in beliefs also translates into differences in employment decisions. However, it is less pronounced, since the men did not follow their beliefs as much as the women did.[27] In Table C1, we investigate the hiring decisions of female and male employers separately.

Table C1: Employment decisions by gender of employer

Share of female candidates chosen		
	(1) Women	(2) Men
Info (initial decision)	0.045	-0.077**
	(0.061)	(0.030)
Info (after information)	0.073	-0.076
	(0.098)	(0.081)
Constant	0.459***	0.613***
	(0.050)	(0.027)
Observations	898	824
Number of employers	102	92

Notes: Results based on a linear probability model taking into account the panel structure of the data using employer random effects. The control treatment is the baseline. Same-gender pairs of candidates are excluded from the analysis. Numbers in parentheses are robust standard errors clustered by session. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

We find that female employers in the control treatment chose the female candidate in 46% of the decisions, on average, while male employers chose the female candidate in 61% of the decisions ($p = 0.002$, Chow test). The gender differences in hiring decisions are smaller in the information treatment. Before receiving information about the specific pair of candidates, men chose the female candidate in 53% of the decisions while women

[27] The share of employers believing that women work more may not be directly comparable to the share of decisions in which an employer chose the female candidate. The reason is that our binary measure of beliefs about who worked more does not take into account the possibility of employers' being indifferent or just having a slight preference for women. However, we believe that it is unlikely that this indifference differs by gender, and can therefore explain the differences between employment decisions and beliefs we observe only for men.

chose the female candidate in 51% of the decision ($p = 0.363$, Wald test). This difference is smaller still if we consider decisions made after receiving information about the specific pair of candidates (the probability that the female candidate would be chosen is 54% for men and 53% for women ($p = 0.841$, Wald test).[28]

[28]In addition, there is no gender difference in updating behavior in the info treatment. Women chose the female candidate if they received a female signal with a probability of 83% and the male candidate after a male signal with a probability of 88%. Men followed a female signal in 84% of the cases and a male signal in 85% of the cases.

D. Instructions

<u>General explanation</u>

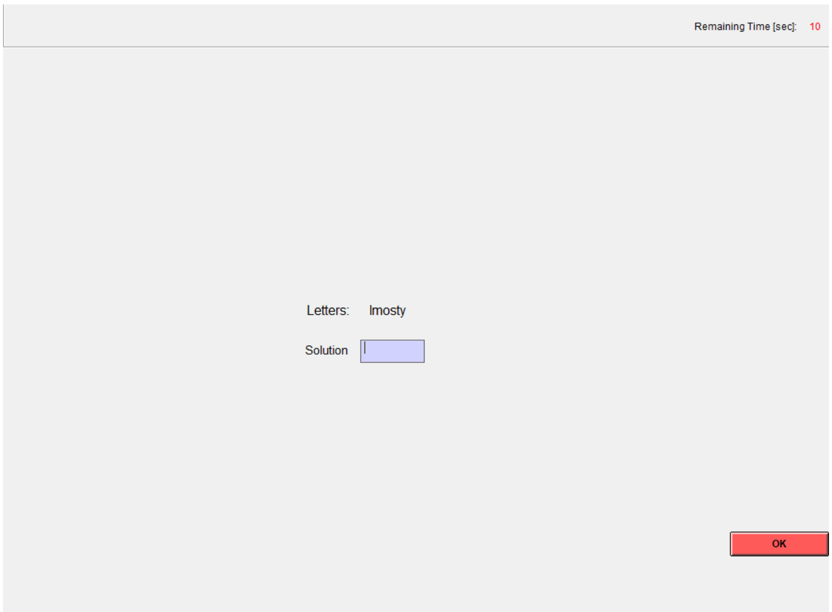
Welcome to our experiment. In this experiment, you can earn money depending on the decisions you make. Therefore it is important that you read the following instructions carefully.

If you have any questions, please raise your hand and an assistant will help you. The use of mobile phones or any other electronic devices is forbidden during the experiment. Do not communicate with other participants during the experiment! Non-compliance will lead to exclusion of the experiment without payment.

The experiment consists of several phases.

Explanation for phase 1

Phase 1 lasts for 5 minutes. You will be asked to arrange strings of six letters in such a way that they form an English word. For example, you could see the letters **"lmosty"** (see example below). By rearranging the letters, you can form the word **mostly.** For each string that you rearrange successfully into an English word, you receive a piece rate of 10 cents.



You can also skip a word, if you are not able to find the solution. However, the maximal number of words you can see is limited to 30.

Explanation for phase 2

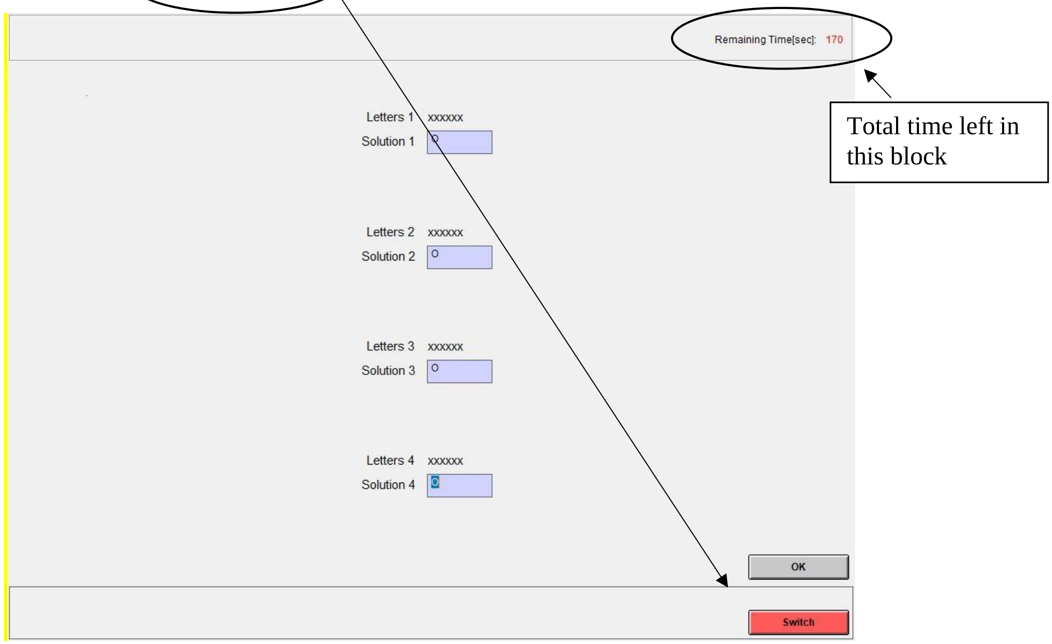
Phase 2 consists of two blocks each lasting 10 minutes. Between the two blocks, there will be a one minute break.

As in phase 1, you will be asked to solve word puzzles. However, now you will see 4 strings of letters on each screen (see example below) and you will only have 60 seconds to find all four words. After 60 seconds, the screen disappears and you have to press the "OK" button to see the next set of words. Be aware that the strings of letters have changed but your previous answers are still in the answer fields. You have to replace the old answers with answers matching the new strings of letters. Every time you manage to solve all four words on a screen correctly within 60 seconds, you receive a bonus of **3 Euros**.

IMPORTANT: You only earn money from word puzzles when you solve all words correctly and press the "OK" button within 60 seconds. You do not earn money from the word puzzles unless you solve all four words on a screen.

As in phase 1, you can also skip a screen, if you are not able to find the solution. However, the maximal number of screens you can see is limited to 20.

In contrast to phase 1, in phase 2 you can also take a break from solving the word puzzles. To do so, you can press the "SWITCH"-button



If you press this button, a new window opens and you can surf the internet. For surfing the internet, you receive 1 cent per second. You can return to the word puzzles, by closing the internet browser and pressing the "back to the word task"- button (see below). Be aware that the strings keep changing every 60 seconds even if you are surfing in the internet.

Important: While surfing the internet you are **not** allowed to try to find answers for the word puzzles. Non-compliance will lead to exclusion of the experiment without payment.

To sum up, you can earn money in this phase in two ways:

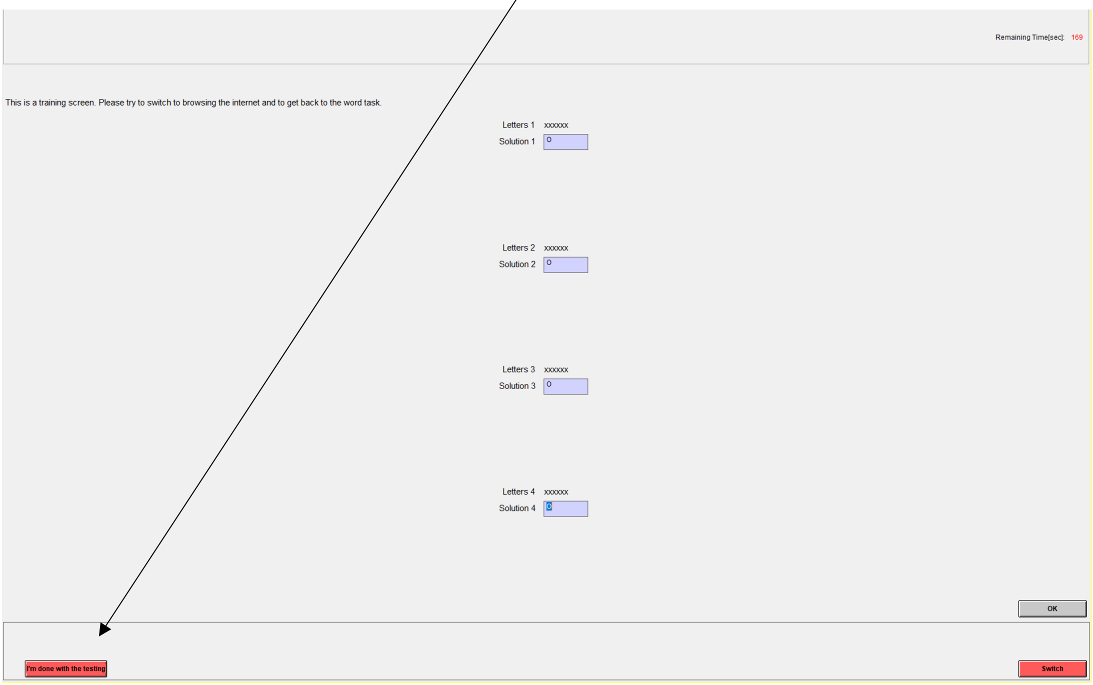
- Every time you solve all four words correctly and press the "OK"- button within 60 seconds, you receive a bonus of 3 Euro.
- If you surf the internet instead, you receive 1 cent per second. (You can switch back and forth between the word puzzles and surfing the internet as many times as you want.)

Example how payoffs are calculated:

Assume that a participant manages two solve 2 complete screens correctly and surfs the internet for 100 seconds. Then he or she earns 2*3=6 Euros as a bonus for solving the word puzzles and 100*0.01=1 € for surfing the internet. This means that the participants earns 7 (6+1) Euros in total.

Before phase 2 starts, we kindly ask you to guess, how many screens you think you will be able to solve correctly.

After that, we will show you a training screen. Please, try switching to surfing the internet and returning to the word task at least one time. After you have switched and returned to the word task, a new button appears in the left button corner of the screen, saying: "I'm done with the testing". Please press this button to proceed to the main part of phase 2.

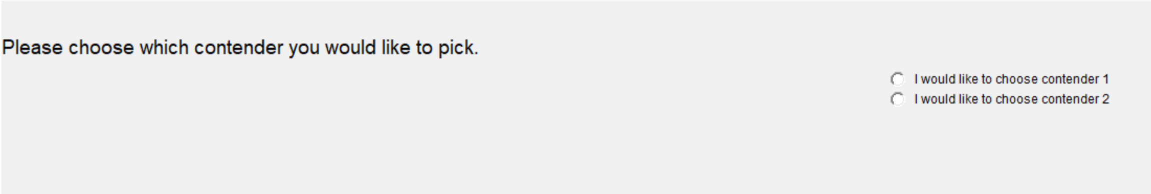


Explanation for phase 3

Phase 3 consists of several, identical rounds. One of these rounds will be randomly selected to determine your payment from phase 3. At the beginning of every round, two participants in the room will be selected by the computer. We will refer to these two participants as *contenders*. We will refer to the rest of you as *observers*. Each participant gets to be a contender at most once during the experiment. In every round, contenders of that round will be asked to stand up and state their numbers, which are assigned according to their cubicles.

Observers

In each round, observers pick one of the two contenders. In the example screen below, contenders have been assigned numbers 1 and 2.


Please choose which contender you would like to pick.

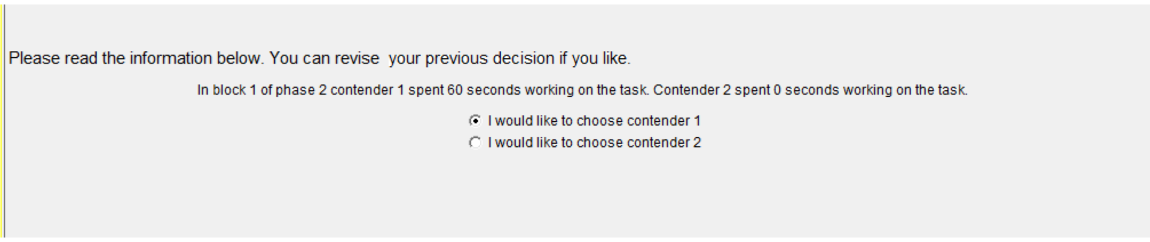
- I would like to choose contender 1
- I would like to choose contender 2

Recall that in phase 2, you participated in two blocks of ten minutes in which the task was to solve all four word puzzles on each screen (in particular, a bonus was paid if a participant managed to correctly solve all four words on a screen within 60 seconds).
Your earnings in Phase 3 depends on your chosen contender's performance and behavior in the **block 2 of phase 2** as follows:

- You receive a bonus of 3 Euros if you pick the contender who spent more time working on the task in block 2 of phase 2.
- You receive a bonus of 1 Euro for each screen your chosen contender completed within the 60 seconds in block 2 of phase 2.

After you pick a contender, we will tell you how much time each of the contenders spent working on the task in block 1 of Phase 2 (recall however that *only* block 2 of phase 2 determines your payment). You can then revise your decision if you like (but you do not have to). Once you make your decision, you must press OK to continue.

For the example screen below, assume that you picked contender 1 before receiving the information regarding how much time each of the contenders spent working on the task in block 1 of phase 2. The screen below gives this information and allows you to revise your decision if you would like to.


Please read the information below. You can revise your previous decision if you like.

In block 1 of phase 2 contender 1 spent 60 seconds working on the task. Contender 2 spent 0 seconds working on the task.

- I would like to choose contender 1
- I would like to choose contender 2

Example of how your earnings are calculated:
Assume that in the round that was chosen for payment, contender 1 spent 500 seconds working on the task and contender 2 spent 100 seconds working on the task in block 2 of phase 2. Furthermore, assume that contender 1 was able to complete 2 screens and contender 2 was able to complete 1 screen in block 2 of phase 2. If you choose contender 1, you receive a **total of 5 Euros** (because you receive 3 Euros from choosing the contender who spent more time working in block 2 of phase 2, and you receive 2 Euros as contender 1 completed 2 screens). If you choose contender 2, you earn **a total of 1 Euro** (because you picked the contender who spent less time working, and contender 2 completed 1 screen).

IMPORTANT REMINDER: The information we give you after you pick a contender is about how much time each of the contenders spent working on the task in **block 1 of phase 2**, but only **block 2 of phase 2** determines your payment.

Explanation for phase 4

In the beginning of phase 4, we ask you to make guesses concerning the performance of the other participants in the room. If your guess is correct, you receive 1 Euro.

After that, you will be asked to make ten decisions between two options, A and B, as shown in the example below.

	Option A	Option B	Your Choice
Decision 1:	Heads: You get 5 points. Tails:You get 3 points.	Heads: You get 7 points. Tails:You get 1 point.	◯ A ◯ B

Please notice that you earn points in this part!

5 points are 1 Euro.

You will only get paid for one of your decisions. At the end of the experiment, we will ask one of the participants to draw a number from 1-10, to determine which of the decisions will be paid out. Furthermore, we will toss a coin publicly to get Heads or Tails. This means that Heads and Tails are equally likely.

For example: Assume decision 1 has been picked to determine your payoff and you decided on option A. Then you have a 50% chance that the coin shows Heads and you receive 5 points and a 50% chance that it will show Tails and you receive 3 points.

The experiment will start soon. Please raise your hand if you have any questions.