Hiring for others:

The role of intermediaries in discrimination*

Link to newest version

Alejandro Hirmas[1] and Jan Hausfeld[1]

[1]University of Amsterdam, Center for Experimental Economics and political Decision making (CREED), and the Tinbergen Institute

November 7, 2024

Abstract

In many hiring processes, job candidates are evaluated by intermediaries, such as human-resources personnel and/or external recruiters. These intermediaries evaluate the candidates based on CVs and other information, and preselect the candidates which they expect to be hired by the hiring manager. As a result, intermediaries might preselect a biased pool of candidates if they are influenced by their expectations about the hiring manager's preferences, potentially including (unwarranted) discriminatory biases. We designed two incentivized experiments, where participants act as intermediaries and predict how a hiring manager will evaluate candidates based on both job-relevant measures (e.g., aptitude and personality tests) and seemingly irrelevant factors such as gender. Importantly, they also observe information about the hiring manager including their gender, age and math skills, allowing us to test whether intermediaries differentially evaluate candidates depending on who they are hiring for. We consistently find that intermediaries expect managers to prefer candidates of the same gender as the manager (expected same-gender favoritism). Furthermore, by tracing the intermediaries' visual attention, we find that intermediaries who expect more same-gender favoritism, also look longer at the candidates' gender information. Accordingly, given the overrepresentation of men in managerial roles, these results on expected same-gender favoritism highlight a novel mechanism in which discrimination takes place in the labor market.

Keywords: Recruitment and Selection, Beliefs, Gender discrimination, Attention
JEL Codes: D81, D83, D87, D91, J7.

*We would like to thank 'A Sustainable Future' (ASF) from the University of Amsterdam for funding this project. We also thank the audience of the CREED seminars, the MPI-CREED workshop, Tilburg University "Attention, Stereotypes and Gender" workshop, and the EGPROC and TiBER conferences.

1 Introduction

In most medium to large companies, recruitment and selection processes heavily rely on intermediaries, such as recruiters and human-resources personnel, to manage the search for new hires. Given that these companies employ more than 50% of the labor force in both the U.S. (McKinsey, 2024) and Europe (Eurostat, 2019), the role of intermediaries in the hiring process is fundamental in shaping the labor force. In the UK, for instance, the average hiring process typically involves at least two rounds of interviews, in which candidates typically meet at least three members of the company (Fenell, 2024), which means that multiple people evaluate the candidate before the hiring manager makes the final decision. The intermediaries need to process the applicant pool to preselect candidates for the hiring manager to choose from. However, who intermediaries believe to be the best candidates may or may not align with the hiring manager's preferences. The goal of this paper is to explore the decision-making process of these intermediaries, focusing on the extent to which their choices are shaped by their beliefs about the hiring manager's preferences.

Internal intermediaries in charge of the recruitment and selection process are usually evaluated and compensated by the effectiveness of the hiring process. Commonly used metrics such as Time to Hire, Interview-to-Offer Ratios and Offer-to-Acceptance rates encourage filling open positions as fast as possible. These metrics are often easier to verify than those promoting long-term outcomes, such as turnover rates, candidate productivity, or candidate satisfaction. Similarly, external intermediaries, such as recruiters or headhunters, are often rewarded only if the candidates they provide are ultimately hired, further strengthening the incentive for all intermediaries to select candidates they believe are most likely to be chosen by the hiring manager. As a result, intermediaries may feel pressured to anticipate the hiring manager's preferences when preselecting candidates. If intermediaries believe the hiring manager to favor certain demographics, such as men over women or younger candidates over older candidates, they may adjust their selections accordingly.

Focusing on gender discrimination, evidence shows that significant biases continue to affect women's opportunities in the workforce. Despite progress in many areas, global female labor force participation remains at just 39% (World Economic Forum, 2023), and 57% of women report experiencing some form of discrimination in their lives including at work (Focus, 2021). Evidence from the U.S. shows that approximately 10% of male managers openly state that they believe men are better suited for leadership roles, compared to only 3% of female managers who hold this belief (Barratt, 2019). Taken together, these figures suggest that biases favoring male candidates remain prevalent. In this paper, we bridge research on intermediaries and gender discrimination, which allows us to answer the following research questions: Do intermediaries anticipate that the hiring manager might hold biased preferences? And if so, do they adjust their own candidates' selection in a way that (unintentionally) reinforces these discriminatory patterns?

We investigate how intermediaries form beliefs about how managers evaluate candidates and how these intermediaries preselect candidates for the hiring manager, who ultimately decides who to hire. Our paradigm represents an agency problem with incomplete information in which the intermediary is not fully aware of the manager's preferences and must act based on their beliefs regarding those preferences. To study the intermediaries' behavior, we conduct two incentivized experiments, in which participants act as intermediaries in a selection process and evaluate candidates based on who they are hiring for. In our first experiment, we elicit how intermediaries expect the manager to predict a candidate's performance based on potentially predictive measures, such as personality and aptitude tests, as well as seemingly irrelevant information like gender. For each candidate, intermediaries must guess the performance predictions of several employers who vary in gender and math skills. Our results show that beliefs regarding predicted performance are strongly influenced by the employers' characteristics. Specifically, intermediaries anticipate that employers will exhibit same-gender favoritism by predicting better performance for candidates of the same gender as the employer, which is unwarranted. Furthermore, intermediaries expect that employers with higher math skills will place more emphasis on candidates' attributes related to math.

In the second experiment, we aim to replicate our findings from the first experiment regarding same-gender favoritism and the differential emphasis depending on the employers' math skills. Additionally, we investigate whether the intermediaries' choices align with the beliefs observed in the first experiment and we also measure visual attention while intermediaries decide. In this experiment, participants preselect one of two candidates based on the same type of information as in the previous experiment. The selected candidate is forwarded to an employer with specific characteristics (gender, math skills, and age). The employer will compare the selected candidate with another candidate from a different source, but similar performance level. If the employer chooses the candidate preselected by the intermediary, the intermediary receives a bonus. To further understand how intermediaries weigh different candidate attributes, we measure how participants allocate their visual attention to these attributes, and use this data to predict the attribute's importance in the decision-making process. We find that intermediaries' choices are consistent with those of the first experiment. Specifically, intermediaries are more likely to select candidates of the same gender as the employer and give greater weight to math-related attributes when hiring for high math-skilled employers. Moreover, our analysis of attention data reveals that participants who consistently allocate more attention to a specific attribute, compared to the rest of the sample, are also more likely to prioritize that attribute in their decision process. This trend is especially pronounced for seemingly irrelevant attributes, such as gender, where participants who focus more on gender are more likely to select candidates of the same gender as the employer. This strong correlation between attention to gender and the expected same-gender favoritism shows that only intermediaries that consistently focus on gender more than the rest of the sample will expect same-gender favoritism.

The results of this paper demonstrate that intermediaries in the selection process are influenced by who they are hiring for. Specifically, intermediaries preselect candidates differently based on who makes the final hiring decision, even if this results in discrimination against certain candidates based on gender. In the current economy, most higher management positions are held by men. If intermediaries expect and act upon the belief that managers prefer candidates of their own gender, then women are less likely to be selected in the hiring process. These findings highlight the value of policies that promote balanced hiring committees to mitigate discriminatory practices stemming from misaligned beliefs. Our attention analysis reveals a significant heterogeneity in how intermediaries aggregate candidate information. These results suggest that implementing clearer selection criteria could be beneficial in improving hiring processes and reducing discrimination.

2 Literature

Discrimination has been widely studied in the field (e.g., Becker, 2010; Bertrand and Duflo, 2017) and in laboratory experiments (e.g., Lane, 2016). Economic theory commonly distinguishes between two types of discrimination, taste-based (Becker, 2010) and statistical discrimination (Arrow, 1973). On the one hand, taste-based discrimination implies that the decision-maker has a clear preference for one group over the other. On the other hand, statistical discrimination arises when the decision-maker has a belief that one group outperforms another group, thus in the absence of more information, will choose based on the group. Although regulatory advances in modern society have made it more difficult for companies to "openly" discriminate, evidence shows that discrimination is still present. Discrimination might come around in more nuanced ways and is even more pervasive (Bartlett, 2009). When comparing labor market outcomes for women and men, women still face higher unemployment rates, worse working conditions and lower incomes (World Economic Forum, 2023).

Another aspect that can play a larger role in hiring decisions is stereotyping. We understand stereotyping as an over-generalization of the characteristics of a group, reducing within-group variance, while increasing between-group variance (Taylor et al., 1978). The role of stereotyping in the labor market has been widely studied. For example, female representation is above 50% for the education and consumer services sectors, while it is significantly lower for the financial and more industrial sectors, such as manufacturing, agriculture and construction (World Economic Forum, 2023). Research shows that women are evaluated worse in male-stereotyped fields, such as information and technology (e.g., Feld et al., 2022). Conversely, women are also preferred in female-stereotyped fields (e.g., Chan and Wang, 2018).

Bordalo et al. (2016) define stereotyping as a bias in the beliefs about the group characteristics. Namely, stereotyping arises due to an overestimation of the representation of the salient aspects of the group. In general, there is ample evidence that negatively-stereotyped groups are discriminated against: they are hired less often (e.g., Coffman et al., 2021; Neumark et al., 1996), they receive worse work evaluations (e.g., Feld et al., 2022). Bartoš et al. (2016) show that negatively-stereotyped groups receive less attention and are consequently overlooked. Moreover, increased attention towards the stereotypes further enhances their role in hiring decisions (Rice and Barth, 2016).

In parallel, labor outcomes can also be affected by the size of certain groups in a company. Namely, in-group favoritism has been shown to affect both hiring (Casoria et al., 2022) and promotion decisions (Ďuriník et al., 2023). Accordingly, when a group is under-represented in a field or a company, they are more likely to be discriminated against (Hagmann et al., 2022). From the supply side, we also see differential exploitation of the in-group favoritism. Bao and Huang (2023) show that men are more likely to benefit from social ties compared to women.

Although evidence shows a persistent impact of gender discrimination in the labor market, the impact of such discrimination differs along the workers' career. The impact of early career discrimination can have long lasting effects. If future promotion and wage-raise decisions depend on current wages, then early-career discrimination will have a persistent impact in the workers' salary progression (Bohren et al., 2019). Alternatively, due to early-career discrimination, it is possible that women that have managed to overcome these difficulties will show outstanding performance due to the strict(er) criteria they had to meet. When analysing career outcomes over time, research has shown that women are less likely to be hired, but once they are hired, they are more likely to be promoted (Groot and Van Den Brink, 1996; Booth et al., 1998, 2003). Indeed, women in leadership positions are seen as more effective than men (Rosette and Tost, 2010), and significantly better for more senior positions, even though women's evaluations are worse compared to men in early stages of their career (Mengel et al., 2019). In line with previous literature, we study whether there is a gender bias in our hiring study. In addition, we explore whether intermediaries are a contributing factor to discrimination since they might be the ones who expect that there is an ingroup bias, and consequently act up on it.

In this paper, we study discrimination in the context of hiring processes. In a typical hiring process, job candidates will most likely be evaluated by several people in multiple stages. In the United Kingdom, the hiring process consists, on average, of two interview stages, while some sectors can have up to six interview stages (Fenell, 2024). The hiring manager will most likely only partake in the last interview stage, leaving the earlier stages to other company members or headhunters. Thus, the intermediaries in the selection process need to take into consideration that someone else is going to make the final decision in hiring. The literature of intermediaries in economic decisions show that both, intermediaries and employers, attribute negative outcomes to the other (e.g., Hamman et al., 2010; Bartling and Fischbacher, 2012; Drugov et al., 2014).

From the perspective of the intermediaries, evidence shows that they are aware and can learn about what their employers want (Smith and Krajbich, 2023). Moreover, intermediaries are sensitive to who they are making the decision for. Experimental evidence shows that intermediaries in hiring decisions disclose different information to managers depending on the gender of the employer. Nonetheless, it is shown that intermediaries expect a priori that employers have similar preferences to theirs (Rubinstein and Salant, 2016). The literature connecting recruiter behavior and gender finds that recruiters are more likely to call candidates of their same gender for interviews (Asanov and Mavlikeeva, 2023), and referrers are more likely to recommend someone of their same gender (Beugnot and Peterlé, 2020). We add to this literature by testing how the gender of the candidate, intermediary and manager interact.

Another source of the differential assessment for women and men can be the choice of which information is used to make a decision. When assessing women and men, evidence shows that decision-makers are likely to use gender-stereotyped characteristics when evaluating candidates (e.g., Friedmann and Efrat-Treister, 2023; Glick, 1991), and evaluate candidates differently depending on the gender stereotype of the field in which the candidate is evaluated (Brock and

De Haas, 2023).

One possible explanation for this differential usage of information across gender comes from a theoretical perspective. Recruiters and hiring managers need to assess the capabilities of the candidate based on much information. Since the selection processes are usually quite intense and fast paced, recruiters need to process as much information as possible in a short amount of time. Recent research showed that recruiters spend approximately six to seven seconds per curriculum when assessing the fit of a candidate (Fenell, 2024; Stepanova et al., 2022). This type of decisions are fitting with the framework of rational inattention models (e.g., Sims, 2003; Matějka and McKay, 2015), where agents know that gathering information is costly and rationally choose how to look for information. Basically, the theory predicts that when gathering information, decision-makers are more likely to collect information that widely represents a group of options versus option-specific information(Peng and Xiong, 2006). Thus, stereotyped information is more likely to be attended to. Moreover, Caplin and Dean (2015) show that when the relevance of information changes due to different incentives, agents look for information differently to maximize their payoffs. In the context of hiring decisions, this implies that recruiters should adjust which information to collect from the candidates depending on who they are hiring for. This selective sampling of information could then lead to biased decision-making both due to the belief about the employer's preference, but also due to biased information processed to make a decision. In order to investigate this biased information processing, we will also measure how intermediaries allocate their visual attention and whether this relates to the behavioral biases.

3 Experimental Design

In this section, we describe the design of the two main experiments of our paper, the dataset comprising the candidates' information, and the experiment for the employers which is used to incentivize the main experiments. To give an overview, Table 1 presents the main characteristics of the different experiments. Across all experiments, participants observed information regarding job market candidates that is usually available in real hiring processes. Importantly, the experiments differ in who makes the decision. First, we ran the employer experiment which was needed to incentivize the main studies. In the employer experiment, we elicit the first-order beliefs of participants regarding the candidates' performance and we collect a series of demographics of these participants. In experiment 1 (Beliefs Study from now on), participants had to predict how these candidates were evaluated by participants from the Beliefs Study (we will refer to these previous participants as employers from now on). In experiment 2 (Choice Study from here onward), participants play the role of an intermediary in hiring decisions. Participants had to select a candidate to forward to an employer, who will compare the selected candidate with another one and then "hire" one of the two[1].

In all experiments, participants had to sign an informed consent, followed by instructions, and they could only proceed if they answered some control questions correctly. Then, the main task started and the experiment ended with a questionnaire. All procedures were approved by the Economics and Business Ethics Committee (EBEC) of the University of Amsterdam.

Table 1: Overview of the experiments

Experiment	Dependent variable	Employers' characteristics	Design
Employers	First-order beliefs	-	-
Study 1 (Beliefs)	Second-order beliefs	Gender, Maths skills	Within
Study 2 (Choices)	Choice between two candidates	Gender, Maths skills and Age	Between

The rest of this section is structured as follows. First, we describe the database used for the candidates' information. Then, we describe the employers' experiment used to incentivize the decisions of the Belief and Choice Studies, which are both explained afterwards.

3.1 Candidates

To create a more realistic choice environment and to be able to incentivize choices, we used data from an application process with real job candidates (Kausel et al., 2016). This database contains the application data from 200 candidates that applied to a job opening (travel agent) for a large airline in the United States and who got subsequently hired. In order to avoid any stereotypes related to the job or the country, participants in our experiments are made aware that the data is real and that the candidates got hired for the same job, but they do not know the location nor the kind of work.

The database contains the gender of the candidate, some personality and aptitudes tests of the candidate collected during the selection process, and, importantly, a performance evaluation three months after they got hired (our measure of true job performance). The application data contains (1) the candidate's *Conscientiousness* level from the Big-five (Fiske, 1949); (2) a test of *General Mental Ability* (*GMA*) measuring their skills for understanding and processing information, such

[1]this hiring stems from the employer experiment and is explained in more depth in section 3.2.

as texts, graphs and tables; (3) the score of an unstructured *Interview* held by a manager of the company; and (4) the reported *Gender* of the applicant. Additionally, we created a random number for each participant which is not correlated with their performance (*Random Number* from now on). The purpose of this variable is to contrast the role of the more meaningful characteristics with another variable that has, by definition, no correlation with a candidate's performance.

We converted the original scores of *GMA*, *Conscientiousness* and the *Interview* into a one to five scale. For each score, the scale represents in which quintile the candidate is positioned with respect to the whole sample of candidates, with one being the lowest quintile and five the highest. The gender is represented by the letters W (Woman) and M (Man). The *Random Number* goes from zero to five[2]. We also divided the performance measure into a one-to-five scale. Again, this measure is what employers needed to predict in the employer study.

In Table 2, we show the summary statistics of the data used for the experiment. In total, the database yielded 130 different combinations of the attributes. For each of these combinations of attributes, we randomly selected one person. The table also shows the coefficients of ordered-logit regressions for the attributes predicting performance. These results suggest that performance is only predicted by *GMA* (p=0.035) and *Conscientiousness* (p=0.041). These coefficients are consistent with the results from the original study, thereby affirming that converting the attributes to quintiles does not significantly affect the attributes' predictive power of performance.

Table 2: Summary Statistics and Regression Results

	N	**Mean**	**SD**	**Min**	**Max**	β	SD(β)
Performance	130	2.89	1.37	1	5		
Female	130	0.60	0.49	0	1	0.626	0.329
GMA	130	2.97	1.41	1	5	0.360	0.114
Conscientiousness	130	2.93	1.42	1	5	0.260	0.114
Interview	130	3.37	1.11	2	5	-0.212	0.144
Random Number	130	2.55	1.80	0	5	0.037	0.091

3.2 Employers

The main goal of the employer study is to create a dataset containing different employers predictions of the candidates performance based on the application information. This dataset is needed to run the intermediary experiments (i.e., the Beliefs study and the Choice study). The employer

[2] We originally aimed to have the *Random Number* ranging also from one to five. Due to a coding error, the number zero was also included in the generating process. We noticed this after the first data collection. Hence, we kept it throughout the experiment.

experiment was ran online via Prolific. In total, 100 participants (48 women, mean age=27.5, all from the Netherlands) took part in this experiment. Since the gender of the employers is an fundamental part for the intermediary studies, we excluded the participants that chose not to disclose their gender (N=2). The experiment could only be done on laptops, and participants were asked to use fullscreen for the duration of the experiment. They were also informed that they could not change tabs while doing the experiments. By doing so, they could be excluded from payment. The experiment took approximately 22 minutes and participants received on average £3.67, including £3.45 as participation fees and an average bonus of £0.22, which depended on their choices.

In this experiment, participants were asked to predict the performance score of a candidate (1-5), based on the candidate's application data (*GMA*, *Conscientiousness*, *Interview*, *Gender* and *Random Number*). Figure 1 shows an example of a decision trial. In total, each participant did 65 trials (plus three practice trials), corresponding to half of the sample of the candidates database. At the end of the experiment, one trial would be selected for each participant and if their prediction was correct, they would get a £0.50 bonus payment.

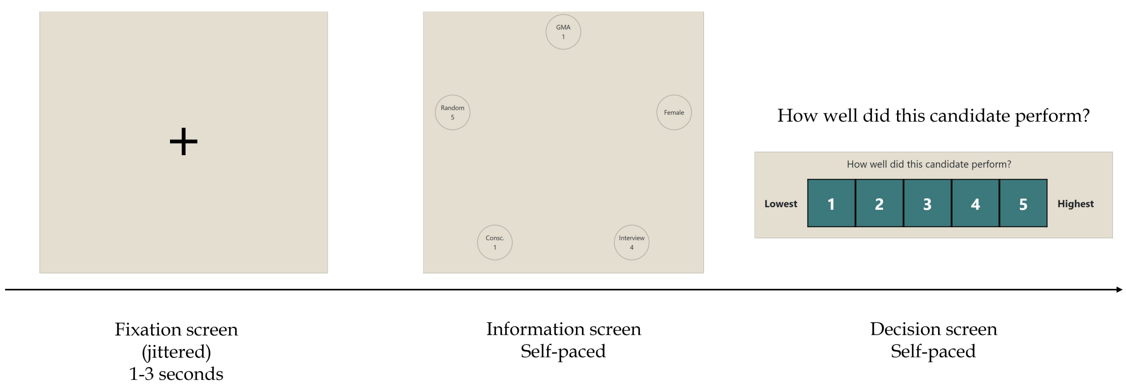

Figure 1: Trial example of the employers experiment

The trial began with a fixation cross for a jittered duration of 1 to 3 seconds. Then, in the information screen, the candidate's attributes (*Gender*, *GMA*, *Conscientousness*, *Interview* and *Random Number*) were located equidistantly from the center and from each other. The order of the attributes was randomized for each participant, but constant throughout the experiment. When ready, they needed to press the 'Enter' key to proceed to the decision screen, where they needed to predict the score of the participant's performance (1-5 scale).

After the main task, participants were asked to fill out a questionnaire with demographics and the Berlin numeracy task (Cokely et al., 2012). The numeracy task consists of four different questions and participants would get £0.10 for each correct answer. Participants also had to indicate their age, gender, socioeconomic status (SES), and their political preferences (left or right). Participants in the Beliefs and Choice studies filled out the same questionnaire at the end of the experiment. Table 3 shows the demographic characteristics of the employers and the participants of the Beliefs and Choice studies. Participants in the three studies were relatively similar, except for the age of participants in the Choice Study which had a few more participants (N=25) with ages above 60[3].

Table 3: Demographics of the multiple studies

	Employers		Beliefs Study		Choice Study	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Berlin-Numeracy Score	2.23	1.12	1.77	1.18	1.82	1.17
Age	27.47	6.47	26.84	7.05	40.11	12.01
SES (1-lowest, 10-highest)	6.30	1.56	5.44	1.62	5.49	1.59
Political (1-Left, 10-right)	4.08	1.97	4.24	1.88	4.25	1.76
N	98		200		411	

3.3 Experiment 1: Beliefs study

The goal of this study is to identify whether intermediaries expect employers to predict candidates differently depending on the employer's characteristics. In the experiment, participants observe a candidate's attributes and need to guess how six employers predicted the candidate to perform. Importantly, the six employers differed in their gender and math skills.

This experiment was done online via Prolific with 200 participants (100 females, 3 rather not say/other, mean(age)=26.84) from the United Kingdom. The experiment took approximately 24 minutes, and participants were paid on average £3.67, including a £3 participation fee and an average bonus of £0.67 depending on their choices. Participants were prompted to keep a fullscreen and to solely focus on the experiment for its duration. Participants that were detected to change tabs frequently, or for too long (more than three minutes) were excluded (N=20).

The main task consisted of 19 trials (including three practice trials). In each trial, participants needed to guess the predictions of six employers for a single candidate. The six employers were participants from the experiment described in section 3.2 and differed in their stated gender (woman or man), and their math skills (low, medium, high). Across the trials, the candidates changed. Figure 2 shows an example of the decision screen, where participants observed the candidate's attributes (above) and needed to input their predictions for the evaluation of the six different employers (below).

[3]We tested whether excluding these participants changed the results and we found similar results.

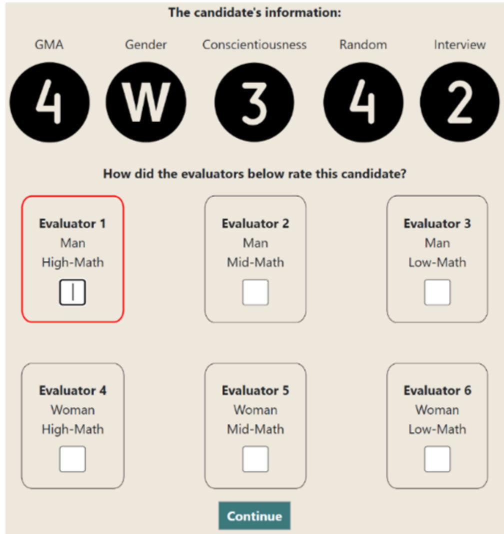

Figure 2: Trial example of Beliefs Study

The information of the candidate is presented in the upper section. *GMA*, *Conscientiousness*, and the *Interview* are presented as 1 to 5 scores. The random number goes from 0 to 5. The gender is presented with a letter W for women and M for men. In the lower part of the screen, participants need to fill the predicted answer for a given employer (performance score from 1 to 5). The order of the employers is pseudo-randomized across participants (see Appendix section B for a more detailed description).

Participants needed to fill in the score (from one to five) they guessed that each employer had assigned to that particular candidate. At the end of the experiment, one trial would be selected and they obtained 0.25 pounds per correct prediction in that trial. The order of the candidate's attributes was randomized for each participant and remained constant throughout the experiment. The presentation order of the employers was also pseudo-randomized at an individual level. The employers in each row always had the same gender, with the gender at the top being random. From left to right, the level of math skills was randomly sorted in an increasing or decreasing order.

3.4 Experiment 2: Choice Study

This study mimics more closely the selection process of a company that is hiring. Participants act as intermediaries and have to select candidates to forward to an employer that will make the final hiring decision. If the candidate is hired, the intermediaries are rewarded. Therefore, it is in their interest to accurately predict how the employer will make the decision. This setup allows us to study whether intermediaries discriminate candidates based on the characteristics of the employer, but this study is about choices while the Beliefs study elicits beliefs.

The experiment took place online via the platform Prolific. We recruited 411 participants (206 Women, 3 "Rather Not Say"/Other, mean(age)=40.11) from the United Kingdom. The experiment took approximately 25 minutes, and participants were paid on average £5.3 on average, including a £3.75 participation fee and an average bonus of £1.55 based on their choices. Similar to the Beliefs study, participants were prompted to keep a Fullscreen and to solely focus on the experiment for its duration.

In this experiment, participants were assigned to one type of 'employer', who was described by gender (woman or man), math skills (low and high) and age range (20-30 or 31-40). We distributed the participants into the eight possible types of employers, such that each condition had a similar amount of participants, and was also gender balanced. We asked participants at the end of the experiment to recall the employer they were assigned to. Participants that did not remember the gender, math skills or age range of their employer were excluded from the analysis (N=20). Table 4 shows the distribution of participants per type of employer and the participants' gender.

Table 4: Participants per treatment condition in the Choice study

Employer			Participants			
Gender	**Maths Skills**	**Age**	**Male**	**Female**	**Other/ Not Say**	**Total**
Man	Low	20-30	24	24	0	48
Man	Low	31-40	23	22	1	46
Man	High	20-31	24	33	1	58
Man	High	31-41	24	25	0	49
Woman	Low	20-32	27	19	0	46
Woman	Low	31-42	25	24	0	49
Woman	High	20-33	24	24	0	48
Woman	High	31-43	23	24	0	47
Total			194	195	2	391

To incentivize the participants' choices, we use the decisions of the employers described earlier in section 3.2. In the Choice study, participants decide between two candidates based on their attributes. To make the pairs of candidates comparable, e.g., to not have very different skilled workers, we calculated the average predicted performance by the employers and sorted them into three performance groups (low, medium and high expected performance). Each decision between two candidates will be about candidates from the same performance category.

The intermediary's task is to forward one candidate to the employer. Then, the decisions of one employer are used who matched the assigned employer's characteristics concerning gender, maths skills and age. Then, we compare the employer's predicted performance of the selected candidate with that of another random candidate of the same performance category. If the selected candidate had a prediction equal or higher than the other, the participant received a bonus of 2 pounds. This incentive scheme aims to simulate the incentives that intermediaries have in a real market. Recruiters usually get a bonus when their referred candidates gets hired. Similarly, the HR personnel in charge of the selection process will get better performance evaluations if their referred candidates gets hired.

In this experiment, we also investigate how intermediaries process the application information, and whether this depends on the employer they make the decision for. To measure the attention process during the decision, we used an attribute-wise mouseover method (Hirmas and Engelmann, 2024), which allows to record for how long, how many times and in which order the different attributes were sampled when making a decision[4]. Figure 3 shows an example of a decision trial of this experiment. First, participants are reminded about the characteristics of the employer they are hiring for (Panel 1). Then, they can scan the information of the two candidates and make a decision (Panel 2A,B). Initially, the information of the candidates is hidden. In order to reveal the information, participants need to move their cursor over the desired attribute in order to see the scores of both candidates simultaneously. When the participants have chosen a candidate, they can click on the respective button below the attributes. After their choice, participants rate how confident they are with their choice (7-point Likert scale).

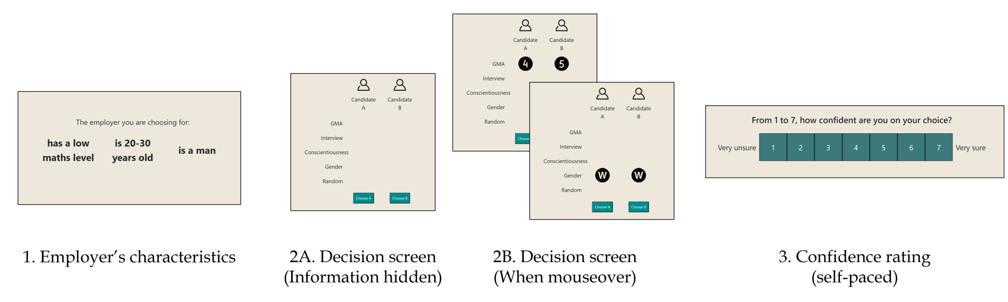

Figure 3: Trial example of the Choice Study

In every trial, participants were reminded of the employer characteristics (Panel 1). Then, participants scan the candidates' information and make a decision (Panel 2A,B). Initially, the information of the candidates is hidden. To reveal the information, participants move their cursor over the desired attribute to see the attribute value for both candidates simultaneously. Participants choose a candidate by clicking on the respective button below the preferred candidate. After choosing, participants rate how confident they are with their choice on a 7-point Likert scale (Panel 3).

Participants chose between 66 pairs of candidates, including 3 practice trials and 9 "obvious" decisions (in the "obvious" decisions, candidates are identical in all dimensions except either one of

[4]We used an attribute-wise mouseover, rather than the typical cell-wise mouseover (MouselabWeb; Willemsen and Johnson, 2011), to reduce cognitive load and decision time, aiming to prevent fatigue and maintain response validity across multiple trials.

GMA, *Conscientiousness* or the *Interview* score, i.e., one candidate weakly dominates the other). The remaining 54 pairs consist of 18 sets of candidates defined by *GMA*, *Conscientiousness* and *Interview*. In these 20 sets, candidates differ in at least two attributes, and since they belong to the same performance groups, their average evaluations do not differ substantially[5]. Hence, these sets of decisions are not trivial and require a proper scanning of all the information. We presented each set three times during the experiment (never one after the other): once where the candidates had the same gender (either two women or two men), and two more times where the candidates differed in gender (man vs woman or woman vs man). This design ensures that if the candidates' gender had no role in the decision, then we should expect similar proportions of men and women selected across the whole sample, but also within each candidate set.

4 Empirical strategy and hypotheses

In this section, we describe our modelling of the decisions of the different agents in the selection process, and introduce our hypotheses for the different studies. Many of our hypotheses across studies are quite similar, hence we compiled them in this section for better clarity. In Appendix section A, we display the original pre-registered hypotheses and how they are linked to the hypotheses defined in this paper.

4.1 Empirical Model

The goal of our empirical analysis is twofold. First, we aim to assess the relevance of different candidate attributes in determining candidate performance, employer predictions, and intermediary beliefs about employer predictions. We model each of these dependent variables as a linear combination of the candidate attributes and estimate a decision weight for each attribute. Second, and most importantly, our goal is to investigate whether the gender of the involved parties (i.e., candidate, employer, and intermediary) influences the decision-making processes of employers and intermediaries. To this end, we will estimate whether there are (expected) biases that arise from specific gender combinations among the relevant parties.

Across the different studies, we estimate a different dependent variable (i.e., performance, beliefs, or choices) based on candidate attributes and the characteristics of employers and intermediaries. Table 5 presents a summary of the empirical models used across the different studies.

[5]For further details regarding the composition of these sets, see Appendix section B.

In the Candidates Dataset, Employers' Experiment, and Beliefs Study, the variable of interest is measured on a one-to-five scale. For these models, we employ an ordered-logit regression. In the Choice Study, however, participants make a binary choice between two candidates, for which use a logistic regression with the selection of the candidate on the right as the dependent variable. In this regression, all explanatory variables are expressed as the difference in values between the right and the left candidate.[6]

Table 5: Model's Regression summary

Dataset	**Dependent Variable**	**Regression Model**	**Gender-related Variables**
Candidates	Performance (1-5)	Ordered-Logit	Female
Employer	1st-order Beliefs (1-5)	Ordered-Logit (RE)	Female Candidate, Match Gender Employer
Beliefs Study	2nd-order Beliefs (1-5)	Ordered-Logit (RE)	Female Candidate, Match Gender Employer, Match Gender Intermediary, Match Gender Both
Choice Study	Choice of Candidate	Logit (RE)	Female Candidate, Match Gender Employer, Match Gender Intermediary, Match Gender Both

All models include the candidates' *GMA*, *Conscientiousness* and *Interview* Scores, alongside the *Random Number*. The Logit model uses choosing the candidate from the right as a dependent variable. All covariates included will be incorporated as differences between the values for the right minus the left candidate. *RE*: Participant Random Effects; *Female Candidate*: 1 if the candidate is a woman and 0 otherwise; *Match Gender Employer*: 1 if the gender of the employer and candidate are the same and 0 otherwise; *Match Gender Intermediary*: 1 if the gender of the intermediary and candidate are the same and 0 otherwise; *Match Gender Both*: 1 if the intermediary, employer and candidate have the same gender, and 0 otherwise

Both the ordered-logit and logit regressions assume that the discrete dependent variable is driven by an underlying latent variable, which we express as a linear combination of the explanatory variables. Let p_c, $B_e(p_c)$ and $B_i^e(p_c)$ denote the latent variables for the performance of candidate c, the beliefs of employer e regarding that candidate's performance, and intermediary i's second-order belief regarding employer e's prediction, respectively. These three latent variables are defined as follows:

[6] This also implies that differences of binary variables can take values of 1, 0, or -1.

$$p_c = \sum_{k \in K} \beta_k^P x_{c,k} + \beta_F^P \mathbb{1}\{g_c = \text{Female}\} \tag{1}$$

$$B_e(p_c) = \sum_{k \in K} \beta_k^{FOB} x_{c,k} + \beta_F^{FOB} \mathbb{1}\{g_c = \text{Female}\} + \beta_{ME}^{FOB} \mathbb{1}\{g_c = g_e\} \tag{2}$$

$$\begin{aligned} B_i^e(p_c) = & \sum_{k \in K} \beta_{e,k}^{SOB} x_{c,k} + \beta_F^{SOB} \mathbb{1}\{g_c = \text{Female}\} + \beta_{ME}^{SOB} \mathbb{1}\{g_c = g_e\} \\ & + \beta_{MI}^{SOB} \mathbb{1}\{g_c = g_i\} + \beta_{MB}^{SOB} \mathbb{1}\{g_c = g_e = g_i\} \end{aligned} \tag{3}$$

Where $x_{c,k}$ represents the value of attribute k for candidate c with $k \in \{$*GMA*, *Conscientiousness*, *Interview*, *Random Number*$\}$. The variables g_c, g_e, g_i denote the genders of the candidate, employer and intermediary, respectively. The coefficients β_F^X, β_{ME}^X and β_{MI}^X, β_{MB}^X represent differences for the performance ($X = P$), first- ($X = FOB$) or second-order beliefs ($X = SOB$) based on whether the candidate is female (F), shares the same gender as the employer (ME), the intermediary (MI), or both (MB).

In the Choice Study, participants compare two candidates and choose which one to forward to the employer. To define the latent variable for this model, we use the same notation as in equation 3, where Δx represents the differences in the variable x between candidates, calculated as the value for the right candidate minus the value for the left candidate:

$$\begin{aligned} \Delta B_i^e(p_c) = & \sum_{k \in K} \beta_{e,k}^{SOB} \Delta x_{c,k} + \beta_F^{SOB} \Delta \mathbb{1}\{g_c = \text{Female}\} + \beta_{ME}^{SOB} \Delta \mathbb{1}\{g_c = g_e\} \\ & + \beta_{MI}^{SOB} \Delta \mathbb{1}\{g_c = g_i\} + \beta_{MB}^{SOB} \Delta \mathbb{1}\{g_c = g_e = g_i\} \end{aligned} \tag{4}$$

The estimated coefficients in all of the above regressions will be proportional to the variance of the latent variable, which means that the scale of the parameters across regressions will vary. To be able to compare these parameters across experiments, we will normalize these coefficients as follows:

$$\omega_k^X = \frac{\beta_k^X}{\sum\limits_{j \in K} \beta_j^X} \qquad \forall k \in K \quad \forall X \in \{P, FOB, SOB\} \tag{5}$$

$$\delta_k^X = \frac{\beta_k^X}{\sum\limits_{j \in K} \beta_j^X} \qquad \forall k \in \{F, ME, MI, MB\} \quad \forall X \in \{P, FOB, SOB\} \tag{6}$$

In this context, the coefficients ω_k^X sum up to one across all conditions, representing a decision weight. Since all attributes are measured in a similar scale (five-point scale), the coefficients δ_k^X will represent variations of the latent variable on that same scale.

4.2 Hypotheses

So far, we have established the empirical models that will be used for our analysis. In this section, we describe the hypotheses that we will test across studies. The focus of this paper is to study the intermediaries' decision process, therefore, our pre-registered hypotheses only describe the analysis of the intermediaries' behavior (Beliefs and Choice Studies), defined by equations (3) and (4) respectively. Nonetheless, we will estimate the models for the candidates' dataset and the employers' experiment to contrast the decision process of the intermediaries with what the actual employers did, and how the candidates really performed. We will divide our analysis into three parts. First, we will analyze whether the gender of the candidates, employers and intermediaries affect the beliefs and choices of the participants. Second, we will see if the intermediaries decisions are also affected by the other characteristics of the employers (math skills and age). Finally, we will incorporate how the intermediaries process the information (i.e., visual attention) and see how the processing of information affects the decision process.

Expected gender biases

The goal of our two studies is to determine if the intermediaries (1) expect employers to have biased preferences based on the gender of the candidates, (2) believe different employers (based on their gender and other characteristics) display different gender biases, and (3) choices are affected by these different expected biases.

We are particularly interested in two potential biases that can be expected from the employers. First, we follow up on the empirical evidence that shows that women still face discrimination in the labor market (World Economic Forum, 2023). In our experiment, we present a gender-balanced sample to participants (i.e., no self-selection), we do not explain the nature of the job and we present equally capable candidates of both genders based on their other attributes (i.e., no reasons for stereotyping nor statistical discrimination). Therefore, if the intermediaries do not expect employers to have discriminatory preferences, female and male candidates should be rated similarly. Since we expect intermediaries to believe and account for discriminatory practices, our first preregistered hypothesis is that the intermediaries will expect discrimination against female candidates.

Hypothesis 1. *Intermediaries expect employers to predict the performance of female candidates as lower than that of male candidates.*

$$\delta_F^{SOB} < 0 \tag{7}$$

Even though we preregistered that intermediaries will expect female candidates to receive lower predictions, we expectdiscrimination to vary across employers. In particular, we are investigating whether intermediaries expect employers to favor candidates of their same gender. This phenomenon, known in related scenarios as same-gender favoritism (e.g., Asanov and Mavlikeeva, 2023; Feld et al., 2022), predicts that employers will favor candidates of their same gender compared to the rest.

Hypothesis 2. *Intermediaries expect employers favorably assess candidates with the same gender as the employer.*

$$\delta_{ME}^{SOB} > 0 \tag{8}$$

The task of the intermediaries is inherently complex, as they must account for potential employer behavior when assessing or selecting candidates. Assuming Hypothesis 2 holds, we expect that intermediaries may have different beliefs about the likelihood of employers engaging in same-gender favoritism. Specifically, if intermediaries anticipate same-gender favoritism from employers and view this as an undesirable behavior, they may be concerned about how such beliefs reflect on their own gender, perhaps even perceiving these beliefs as an indirect judgment of themselves. Alternatively, if participants feel more comfortable predicting the behavior of an employer of their same gender, it is possible that they stereotype the employer of a different gender by expecting that they will engage more in same-gender favoritism. Therefore, we hypothesize that intermediaries will predict stronger same-gender favoritism from employers of the opposite gender.

Hypothesis 3. *Intermediaries expect stronger same-gender favoritism from employers of a different gender than their own.*

$$sign(\delta_{ME}^{SOB})\delta_{MB}^{SOB} < 0 \tag{9}$$

Equation 9 indicates that the effect of a candidate matching both the employer's and intermediary's gender will differ in direction from the effect of matching only the gender of the employer. This implies that intermediaries expect lesser same-gender favoritism coming from employers who share their gender compared to employers of a different gender.

The role of the employers' characteristics

In both, the Beliefs and Choice study, the employers are described not only by their gender, but also by their math skills (and also age in the Choice Study). The goal of introducing the math skills is to see if intermediaries also engage with other employer's characteristics when evaluating the candidates. Specifically, we conjecture that intermediaries will prioritize candidates' analytical skills when hiring for employers with high math skills, and they will rather prioritize communication skills and personality when hiring for employers with low math skills.

Hypothesis 4. *Intermediaries expect different decision weights for employers with higher (compared to lower) math skills.*

$$\omega_{e,k}^{SOB} \neq \omega_{e',k}^{SOB} \wedge \delta_{e,k}^{SOB} \neq \delta_{e',k}^{SOB} \qquad \forall k \quad \forall \{e, e' | M_e \neq M_{e'}\} \tag{10}$$

Where $\omega_{e,k}^{SOB}$ represents the expected decision weight for attribute k for a given employer e (as in equation 6), and M_e represents the employer's math skills. To test this hypothesis, we include a moderating effect of the employers' math skills into our equations (3) and (4). Namely, we define the attributes coefficients as:

$$\beta_{e,k} = \beta_k + \beta_{MM,k}\mathbb{1}\{M_e = \text{Medium}\} + \beta_{HM,k}\mathbb{1}\{M_e = \text{High}\} \tag{11}$$

Where M_e represents the employer's math skills (low, medium or high). It must be noted that in the Choice study, the employers' only had low and high math skills. The introduction of the employers' age range in the Choice study allows to test if more dimensions decrease or even eliminate the role of the employers' gender and math skills on the decision. Therefore, we have no pre-registered hypotheses regarding the employers' age. Nonetheless, we aim to do a similar (exploratory) analysis to study the moderating effect of the employers' age on the decision weights and expected biases.

The role of attention in the decision process

In the Choice study, we measure how participants sample the candidates' information when making a decision by measuring visual attention. Research on visual attention in decision-making has established a strong correlation between the importance of information in a decision and the visual attention allocated to that information (Orquin and Mueller Loose, 2013). Specifically, the literature on gender discrimination that incorporates visual attention finds evidence that individuals who discriminate search the available information differently than those who do not (e.g., Bartoš et al., 2016; Stepanova et al., 2022). Based on this evidence, we aim to test whether the participants' attention to the different types of candidate information can predict the importance of these attributes in intermediaries' decisions.

Hypothesis 5. *Intermediaries who consistently allocate more visual attention to a specific attribute, relative to the rest of the sample, are more likely to base their decisions on that attribute.*

To test this hypothesis, we first define how we measure the visual attention to an attribute. Following Rahal and Fiedler (2019), we use the proportion of time spent looking at an attribute as a reliable indicator of its relative importance in multi-attribute decisions. Since our goal is to capture individual differences in decision-making processes, we apply the methodology described in Hirmas et al. (2024), which involves constructing an attribute-specific attention index which is incorporated in decision models. This index reflects the relative attention each individual allocates to each attribute compared to the rest of the sample. Specifically, we construct this index by calculating each individual's average attention measure (proportion of time) across all decision trials. Assuming that each individual's decision strategy remains stable throughout the experiment, this average serves as a more reliable indicator of the attribute's relative importance across individuals than trial-specific attention measures.

To analyze whether these attention indices predict differential decision weights and expected biases in the decision, we define the attribute coefficients and individual attention indices as:

$$\beta_{i,k} = \beta_k + \beta_{\bar{a},k} Z(\bar{a}_{i,k}) \qquad \forall k \in K \bigcup \{\text{F,ME,MI,MB}\} \tag{12}$$

$$Z(x) = \frac{x - \bar{x}}{s_x} \tag{13}$$

$$\bar{a}_{i,k} = \sum_{t \in T} a_{i,k,t} \tag{14}$$

Where $a_{i,k,t}$ is the attention allocated by participant i in trial t to attribute k, and $\bar{a}_{i,k}$ represents the average attention allocated by the same participant to that attribute across all trials. We use a standardized version of this individual average, denoted by $Z(\bar{a}_{i,k})$, which allows us to interpret the coefficient $\beta_{\bar{a},k}$ as the change in the attributes' weight when the participant allocates one standard deviation more of attention to that same attribute.

If these individual average attention indices predict differences in the decision weights and expected biases, the next step is to study which factors predict the individual differences in attention. Specifically, we test whether the individual differences in attention can be explained by the employer characteristics (gender, math skills and age).

Hypothesis 6. *The individual-average attention allocated to an attribute depends on the type of employer the individual is making the decision for.*

To test this hypothesis, we run a linear regression on each attribute's individual attention indices using the characteristics of the individual, the employer and the positions of the attributes as predictors[7].

5 Results

We divide the analysis of our studies into three parts. First, we analyze the data of the Beliefs and Choice Studies alongside the candidates and the employers datasets to test which attributes are relevant to the choices, and which gender biases appear in the different studies. Subsequently, we analyze how the employers' math skills affect the intermediaries' decisions. Finally, we test whether intermediaries allocate attention differently depending on the employers attributes, and whether individual differences in the intermediaries attention capture heterogeneous behavior in their choices.

[7]Note that these regressions are at an individual level (i.e., one observation per participant)

5.1 Main Analysis

In the Beliefs and Choice Studies, participants take the role of an intermediary, and need to choose based on how they think the employer is going to rate the candidates. In this section, we estimate the decision models described in Section 4.1. We also present the estimations for the employers' decisions, and the actual candidates performance, in order to contrast i) whether the intermediaries accurately predict the employers choices and ii) and employers accurately predict the candidates' performance.

Table 6 shows the estimates of the regressions for the candidates' performance (column 1), the employers' predictions (column 2), and the intermediaries decisions in the Beliefs and Choice Studies (columns 3 and 4 respectively). As can bee seen in column 1, *GMA* ($p = 0.001$) and *Conscientiousness* ($p = 0.026$) are the only predictors of the candidates' performance. We also do not find strong gender differences when comparing the performance of male and female candidates ($p = 0.065$).

Table 6: Estimations of decision models for different studies

	(1) Candidates	(2) Employers	(3) Beliefs Study	(4) Choice Study
Female Candidate	0.626	0.221**	-0.000	0.102*
	(0.340)	(0.076)	(0.063)	(0.044)
Female Employer		0.262	0.046	-0.025
		(0.207)	(0.025)	(0.043)
Matching gender Employer		-0.159*	0.368***	0.440***
		(0.075)	(0.038)	(0.041)
Matching gender Intermediary			0.104	-0.097
			(0.065)	(0.056)
Matching gender Both			-0.211***	-0.067
			(0.050)	(0.119)
GMA	0.360**	1.580***	0.876***	1.167***
	(0.110)	(0.103)	(0.023)	(0.047)
Conscientiousness	0.260*	1.720***	0.723***	0.800***
	(0.117)	(0.123)	(0.024)	(0.029)
Interview	-0.212	1.065***	0.560***	0.859***
	(0.149)	(0.119)	(0.024)	(0.036)
Random Number	0.037	-0.029	0.087***	0.163***
	(0.103)	(0.023)	(0.018)	(0.012)
Dependent variable	Performance	First-order Beliefs	Second-order Beliefs	Choices
Observations	130	3185	17280	21060
AIC	.	5086.019	41761.207	20731.920
BIC	.	5158.813	41869.810	20819.426
Random-Intercept	No	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In column 2, we find that *GMA* ($p < 0.001$), *Conscientiousness* ($p < 0.001$) and *Interview* ($p < 0.001$) are predictors of the employers' first-order beliefs. Regarding the gender biases, we find that female candidates are rated higher than male candidates ($p = 0.004$) and that candidates of the same gender as the employer are rated lower ($p = 0.036$). When considering both biases jointly, we find no significant difference in female employers' beliefs about male and female candidates ($\beta_F + \beta_{matchE} = .070$, $p = 0.5818$).

When analyzing the decisions of the intermediaries across the Beliefs (BS, column 3) and Choice Studies (CS, column 4), we find that *GMA* (BS: $p < 0.001$, CS: $p < 0.001$), *Conscientiousness* (BS: $p < 0.001$, CS: $p < 0.001$) and *Interview* (BS: $p < 0.001$, CS: $p < 0.001$) are positively correlated with the intermediaries' decisions. Surprisingly, we also find that intermediaries expect employers to use the *Random number* (BS: $p < 0.001$, CS: $p < 0.001$).

Regarding the expected gender biases, we find no expected discrimination against female candidates in the Beliefs Study ($p = 1.000$), but a positive expected bias in favor of female candidates in the Choice Study ($p = 0.072$). Both results contradict our Hypothesis 1 stating that intermediaries expect discrimination against female candidates.

Result 1. *There is no evidence of discrimination against female candidates.*

When analyzing the employers' expected same-gender favoritism, we find that intermediaries expect employers to give higher ratings to candidates who share the employer's gender in both studies(BS: $p < 0.001$, CS: $p < 0.001$). These results support Hypothesis 2 which states that intermediaries expect same-gender favoritism coming from the employers.

Result 2. *Intermediaries expect same-gender favoritism coming from the employers.*

We find no evidence supporting that intermediaries favor candidates matching their own gender (BS: $p = 0.109$, CS: $p = 0.086$). On the other hand, the Beliefs Study shows a negative impact on the candidates' outcomes if they match the gender of both the employer and the intermediary ($p = 0.000$). Taking into account the results regarding expected same-gender favoritism, we interpret this negative interaction effect (matching the gender of both, intermediary and employer) as intermediaries expecting employers of the opposite gender to engage more in same-gender favoritism. Nonetheless, these results are not confirmed by the Choice Study, where we do not find a significant effect of matching both genders ($p = 0.574$). Therefore, we do not find confirmatory evidence for our Hypothesis 3, which predicts that the expectation of the opposite gender to engage more in same-gender favoritism. The supporting results are at most, context-dependent, since this effect appears only when eliciting beliefs and not in their actions.

Result 3. *Intermediaries believe employers of their opposite gender to engage in more same-gender favoritism, but the intermediaries' actions do not reflect such beliefs.*

As mentioned in section 4.1, the parameters of the models from different studies are not directly comparable because they are scaled by the inverse of the variance of each model's error term, which naturally varies across studies. To address this issue, we apply the Delta-method to rescale the parameters, following the approach outlined in equation (6). After rescaling, the decision weights of *GMA*, *Conscientiousness*, *Interview*, and the *Random Number* sum up to one, and the rescaled expected biases now reflect points added or subtracted on the five-point scale measuring performance. Figure 4 shows the decision weights on Panel A, and the (expected) gender biases in

Panel B. When comparing the decision weights, we find that intermediaries are relatively accurate when predicting the relevance of *GMA* and the *Interview* in the employers choices, while they underestimate the influence of *Conscientiousness*. For the gender expected biases, we find an overestimation of the same-gender favoritism which is similar for both, the Beliefs and Choice Studies. Additionally, the biases associated with matching the intermediary's own gender and matching both differ significantly across studies. However, these biases are not directly comparable to those of the employers as the employers' decisions reflect first-order beliefs .

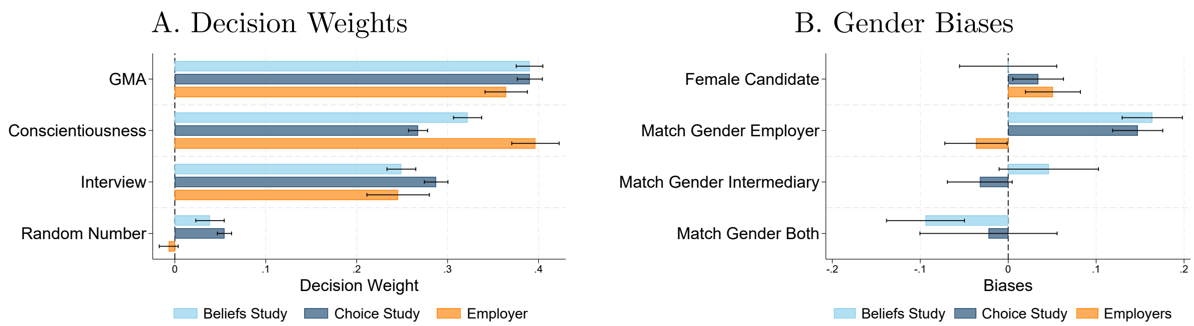

Figure 4: Decision weights and biases
The figures above show the estimated decision weights and biases for the Beliefs Study (light blue) and Choice Study (blue), and the employers database (orange). The weights and biases are estimated based on equation 6 using the Delta-method. The parameters come from the estimates of Table 6 with confidence intervals at 95%-level.

So far, we have established how intermediaries make decisions when considering the employers' gender in their choices. In the next section, we explore whether other characteristics of the employers also play a role in the decision.

5.2 Role of other employers' attributes

In the Beliefs and Choice Studies, intermediaries need to judge a candidate while taking into account the characteristics of the employer they are matched with. In the Beliefs Study, intermediaries observe the gender and math skills of several employers, while in the Choice Study, they are matched with one employer, who is described by their gender, math skills, but also their age. In this section, we analyze whether these other employer characteristics' (math and age) also affect how the intermediaries make their decisions.

In our pre-registered hypotheses (See Appendix section A), we conjectured that intermediaries will expect employers with higher math skills to put more emphasis on analytical skills (*GMA*) compared to the other candidates' attributes (*Interview* and *Conscientiousness*). To test this, we estimated the model presented in the previous section with a moderating effect of the employers' math skills for each attribute (see Appendix section D for the regression tables). In order to compare how the employers' math skills affect the decision process, we use the same approach as in the previous section. For each level of math skills, we rescale the decision weights such that they sum up to one, and the expected biases to reflect points added or substracted to the five-point scale measuring performance.

Figure 5 shows the decision weights (left side) and gender biases (right side) for the Beliefs (first row) and Choice (second row) Studies conditional on the employers' math skills. The results on the decision weights show three consistent patterns across studies. First, intermediaries expect employers to assign a higher weights to *GMA* when choosing for high-math employers as compared to low-math employers (BS: $\Delta(High - Low) = 0.088, p < 0.001$, CS: $\Delta(High - Low) = 0.055, p = 0.001$). Conversely, we find that intermediaries expect a lower weight on the *Interview score* (BS: $\Delta(High - Low) = -0.048, p = 0.002$, CS: $\Delta(High - Low) = -0.027, p = 0.045$) for high-math employers. In both studies, the estimated weight for the *Random Number* is lower for high-math employers, but these differences are not significant (Study 1: $\Delta(High - Low) = -0.028, p = 0.061$, Study 2: $\Delta(High - Low) = -0.014, p = 0.078$).

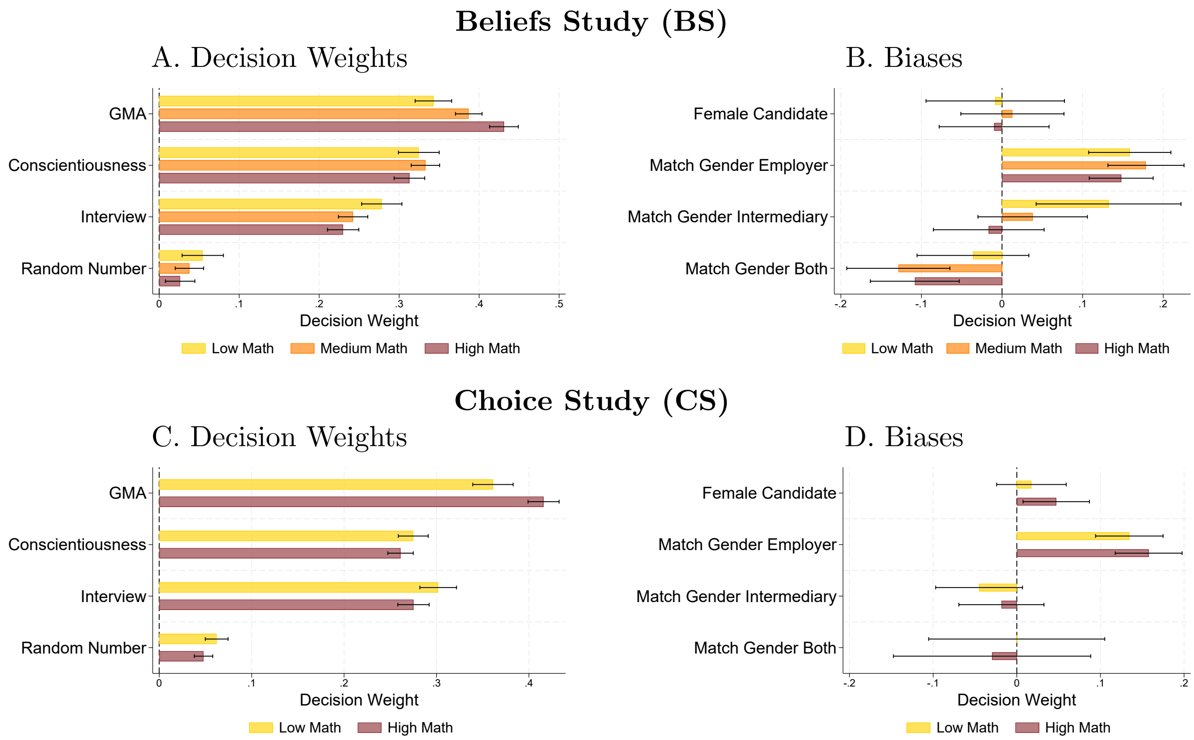

Figure 5: Decision weights and biases based on employers math skills

The figures above show the estimated decision weights (BS in panel A, CS in panel C) and biases (BS in panel B, CS in panel D) conditional on the employer's math skills. The employer's math skills were presented as low (yellow), medium (orange) or high (red) in BS, and low (yellow) or high (red) in CS. The weights and biases are estimated based on equation 6 using the Delta-method. The parameters come from the estimates from Appendix Table 10 with confidence intervals at the 95%-level.

When we compare the differences in the gender biases, both studies show that intermediaries expect same-gender favoritism from the employer, but with no significant differences across math skills. Nonetheless, the joint test of whether all the decision weights and biases differ between low- and high-math-skilled employers is significant in both studies (BS: $\chi^2(7) = 66.22$, $p < 0.001$, CS: $\chi^2(7) = 18.73$, $p = 0.009$). These results support Hypothesis 4, claiming that intermediaries expect employers with different math skills also evaluate candidates differently.

Result 4. *Intermediaries expect high-math-skilled employers to allocate higher decision weights to GMA and lower weights to the Interview when compared to low-math-skilled employers.*

Although, we did not have any pre-registered hypotheses regarding the employers' age range and their expected choices, we conduct an exploratory analysis of whether intermediaries expect differences in the decisions of the employers depending on their age range. The results show no significant changes in the estimates (see Supplementary Table E for the estimations).

5.3 Attention in the decision process

In the Choice Study, we recorded the participants' information search, i.e., which information is looked at and when. In this section, we test (i) whether differences in the intermediaries' decisions are correlated with differences in their attention processes, and (ii) whether intermediaries allocate attention differently for different employers.

To analyze the participants' visual attention allocated to the attributes, we follow the methodology of Hirmas et al. (2024) and the described approach from section 4.2 by calculating individual averages of the proportion of time spent looking at each attribute. These indices capture consistent patterns in participants' information-search processes, reflecting individuals' deviations from the sample's average behavior. We measure visual attention using the proportion of time spent on each attribute, as suggested by Rahal and Fiedler (2019). However, as shown in Appendix I, our analysis produces similar results when using alternative, commonly employed attention measures.

Before analyzing the link between the individual-attention indices and the decision process, we first examine the distribution of these indices for each attribute across the experimental sample. Figure 6 presents a violin plot showing the distribution of participants' average attention to each attribute. Participants allocate similar proportions of time to *GMA*, *Conscientiousness* and *Interview*. Attention to *Gender* is slightly lower and the lowest mean is observed for *Random*. Additionally, the distribution of attention to the random number attribute is more dispersed compared to the other attributes, suggesting greater heterogeneity in how participants process this information.

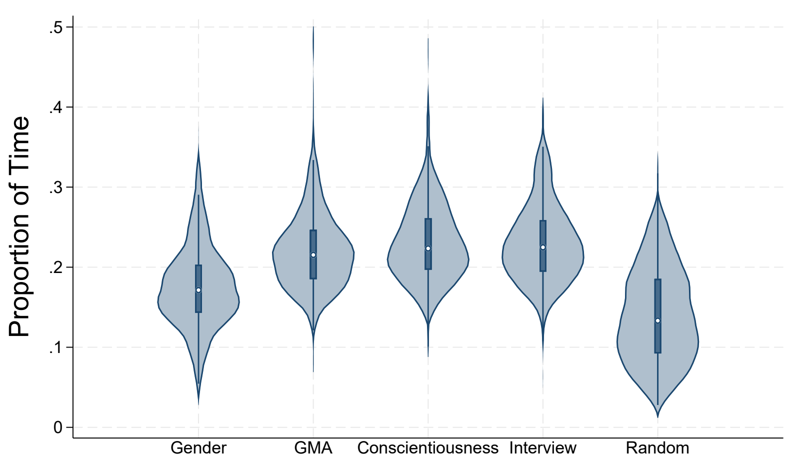

Figure 6: Individual-average attention for each attribute

To analyze whether participants who allocate more attention to an attribute relative to the rest of the sample, also weigh this attribute more heavily, we estimate the decision weights and expected biases using equation (12). Here, the importance of each attribute is moderated by its respective attention index. With this approach, the parameters associated with each attribute vary according to individual attention indices, so we rescale these coefficients at the individual level as well. Since this rescaling involves dividing each parameter by the sum of the parameters for *GMA*, *Conscientiousness*, *Interview*, and the *Random Number*, the coefficients are not only dependent on the attention allocated to each specific attribute but also on the attention allocated to all other attributes.

Figure 7 shows the individual attention-based decision weights and gender biases (vertical axes) compared to the individual-average attention (horizontal axes) for each intermediary. The estimates used to calculate these weights are presented in Appendix section F. To compare the impact of attention to behavioral data, we contrast the individual decision weights and biases with the estimates from the baseline model (horizontal line including 95% confidence intervals). The individual estimates show a large variation for almost all attributes (*GMA*, *Conscientiousness*, *Interview* and *Random Number*), but also for the matching-employers' gender bias when compared to the confidence intervals of our baseline model. Moreover, we find that the relationship between the weights is stronger for matching the employer's gender and for the *Random Number*.

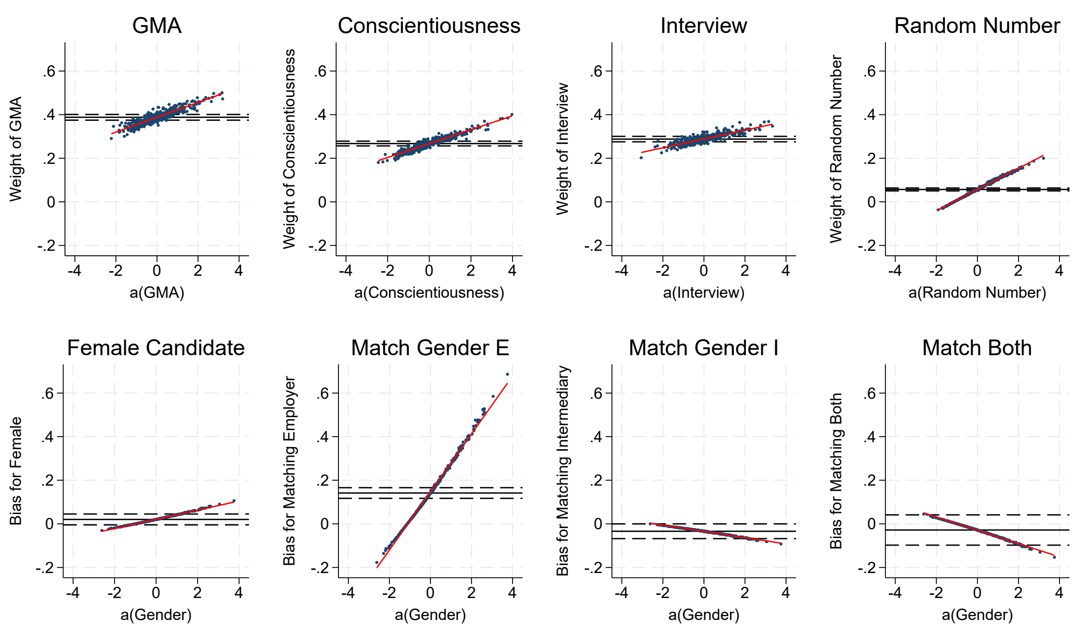

Figure 7: Attention-based decision weights

These results are in line with Hypothesis 5, which predicts a positive correlation between the relative attention to an attribute and its importance in the decision, measured by the decision weights and expected biases. These results are based on our parametric specification, but are also supported by the raw data (see Appendix section G).

Result 5. *Intermediaries that spend more time looking at GMA, Conscientiousness or Random Number, are more likely to choose candidates with higher scores in the given attribute. Similarly, intermediaries spending more time at Gender, are more likely to choose candidates who match the gender of the employer.*

The attention model in Hirmas et al. (2024) identifies two requirements for attention indices to improve model predictions. First, there must be a non-zero correlation between an attribute's importance (decision weight) and the attention allocated to it, which aligns with existing attention literature showing significant correlations between information importance and attention (Orquin and Mueller Loose, 2013). Second, the sample must exhibit heterogeneous preferences (weights) for the attribute, as attention cannot predict variations that not present in the sample. In our experiment, two types of information about candidates are considered, the information that is likely to be relevant, such as *GMA*, *Conscientiousness*, and *Interview*, which serve as proxies for skills essential for most jobs; and the seemingly irrelevant information like the random number, which by construction unrelated to candidate performance. The candidate's gender would only predict performance if it reveals differences not already captured by other attributes. Since participants are unaware of the job type, there is no reason to expect one gender to outperform the other. Thus, gender and random number become relevant only if intermediaries expect employers to consider them, leading to greater heterogeneity in intermediaries' beliefs about these attributes' influence on employers' decisions.

Now that we have established a positive correlation between the systematic differences in attention to the attributes across participants and the importance they allocate to them, we can analyze if there are any factors that predict these differences in attention. Specifically, we test whether intermediaries' attention to certain information is driven by the employer they are hiring for, or by any characteristics inherent to the intermediary. Table 7 shows the results of the regressions of individual average attention for each attribute on the different individual-specific factors. The factors are separated in three categories, employers' characteristics, individual characteristics and the position of the attribute.

Table 7: Drivers of individual differences in attention to the different information

	(1) Gender	(2) GMA	(3) Conscientiousness	(4) Interview	(5) Random
Age(Employer)	0.003	0.002	-0.004	-0.006	0.004
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)
Female(Employer)	0.005	0.005	-0.004	-0.003	-0.003
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)
Math(Employer)	0.002	0.000	-0.003	0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Age(Intermediary)	-0.000	0.000	0.000	0.001**	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female(Intermediary)	-0.001	0.005	0.003	0.000	-0.002
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)
Math(Intermediary)	-0.003	0.006	0.016***	-0.001	-0.015*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)
SES(Intermediary)	0.003	-0.000	-0.001	-0.002	0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Political orientation(Intermediary)	-0.003	0.001	0.002	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Second position	-0.075***	-0.068***	-0.071***	-0.082***	
	(0.007)	(0.006)	(0.006)	(0.006)	
Third position	-0.050***	-0.051***	-0.060***	-0.056***	
	(0.007)	(0.006)	(0.006)	(0.006)	
Fourth position	-0.048***	-0.044***	-0.046***	-0.052***	
	(0.007)	(0.008)	(0.006)	(0.007)	
Constant	0.211***	0.230***	0.277***	0.282***	0.169***
	(0.017)	(0.015)	(0.016)	(0.017)	(0.022)
Observations	390	390	390	390	390
AIC	-1289.433	-1316.971	-1367.945	-1364.908	-1095.113
BIC	-1241.839	-1269.377	-1320.351	-1317.314	-1059.418
Employer variables: F(3,N-k)	0.736	0.493	1.475	0.777	0.248
p-value	0.531	0.687	0.221	0.507	0.863
Intermediary variables: F(5,N-k)	1.553	1.201	3.008	2.005	3.556
p-value	0.173	0.308	0.011	0.077	0.004
Positional variables: F(3,N-k)	36.237	43.310	51.703	74.103	
p-value	0.000	0.000	0.000	0.000	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Across all columns, we do not find any significant differences in attention based on the employers characteristics. On the other hand, the intermediaries characteristics have a significant impact for Conscientiousness, Interview and the Random number. Specifically, we find that intermediaries with higher math skills focus more on *Conscientiousness* and less on the *Random Number*. Furthermore, we also find that older intermediaries focus more on the interview score and less on the random number. Nonetheless, these differences are small in scale. Finally, we find strong effects on the attention the intermediaries allocate to the attribute depending on the position of the attribute. These results reject our Hypothesis 6 stating that the employers' characteristics affect how the intermediaries process the information.

Result 6. *The intermediaries' attention allocated to the different attributes is driven by the position* *of the attribute, and partially correlates with the intermediaries' characteristics.*

6 Discussion

This study investigates how intermediaries may reinforce gender-based discrimination in hiring processes. Using two novel, incentivized experiments, we show that intermediaries believe that employers favor candidates of the same gender. When intermediaries choose which candidates to recommend, they account for this perceived bias by favoring same-gender candidates.

To explore whether intermediaries' expectations of the same-gender favoritism vary across our sample, we analyze the attention intermediaries allocate to different candidate attributes during decision-making. Our findings reveal considerable heterogeneity in attention to attributes, especially those where the relevance is less obvious, such as a candidate's gender or randomly assigned numbers. In contrast, presumably job-relevant attributes, like general mental ability (GMA), are value more consistently across intermediaries. These insights suggest that intermediary biases are not uniform but rather reflect individual perceptions of employer preferences, which can contribute to discriminatory practices in hiring.

While we do find strong differences in behavioral responses for different employers, we do not find any association between the different attention patterns and who the intermediaries where hiring for. On the other hand, the intermediaries' attention was modestly explained by their individual characteristics, specifically their age and maths skills. We further investigated whether the differences in the decision weights captured by the attention indices, can also be captured with other elicited demographics. In Appendix table 12, we show that if we incorporate these other individual characteristics into our attention model, we still find that the individual-average attention significantly moderates the decision weights. Thus, we conclude that the differences in the attention process predict distinct decision rules that could not be explained by other commonly elicited demographic measures.

Our analysis is based on the individuals' average attention, and does not account for any potential intra-trial differences in attention. In Appendix section H, we show that there are no strong differences in attention over time (i.e., trial number), nor across the attributes and gender combinations presented in the different trials. Hence, if there are intra-trial attention effects, these effects are of small magnitude. We also analyze whether our results are dependent on the attention measures used. In Appendix section I, we show the estimates of our attention model when using other commonly used measures such as total dwell times and number of fixations. The results are similar for the dwell times and somewhat weaker for the number of fixations, suggesting that the effects of attention are captured mostly by how long the participants attend to the attribute instead of how many times. Nonetheless, the attention model using proportion of time (as in the results section) has the best model fit corroborating the usage of that measure of attention.

Our attention model assumes that the attributes' weight is only predicted by the attention to that specific attribute, and not the remaining attributes. To explore whether the attention process as a whole predicts the participants decisions even better, we perform a clustering exercise in Appendix section K which is based on the participants' individual-average attention to all attributes. Using k-means clustering, we find that participants are mostly separated into two groups: one group that focuses more on *GMA*, *Conscientiousness* and *Interview* (attribute-focus group from now on) compared to the rest of the sample, and the other group focuses mostly on gender and the random number (gender/random focus from now on). The latter might the group of participants that are also more susceptible to experimental demand, and therefore it is interesting to study both groups separately. For that purpose, we estimate the baseline decision model for both groups and find the following. First, the participants in the attribute-focus group display higher weight for *GMA* and *Conscientiousness* compared to the gender/random focus. The gender/random focus group shows a larger expected bias for candidates with the same gender as the employer and for the random number. Additionally, all the gender biases are significant (including matching intermediary's gender and matching both genders). It is important to note, the attribute-focus group still expects a large and significant effect for matching the employer's gender, while we find no other gender expected biases. These results suggest that even if there are experimental demand effects driving part of these expected biases, the effect of matching the employer's gender is still robust.

In the Choice Study, we randomized the order in which the attributes are presented. We found a strong correlation between the attention spent to the attribute and the position of that attribute. This result raises the question of whether the link between the attention allocated to an attribute and its importance in the decision is mostly or solely driven by the positional effects. In Appendix table 12, we show that if we decompose the individual differences in attention, i.e., separating the individual differences from the positional effects, the attention measures are still significantly correlated with the attributes importance in the decision. Moreover, the position of the attribute does not significantly affect the decision process. These results are in line with the attention literature which states that when decision-makers (e.g., the intermediaries) have strong decision rules, they are not influenced by bottom-up drivers of attention, such as the position of the attribute.

Now that we have established that intermediaries expect same-gender favoritism coming from the employers, and that the intensity of these expectations varies highly across intermediaries, we will run further exploratory analyses contrasting our results with potential mechanisms in which the same-gender favoritism can take place.

The literature has identified several nuanced mechanisms that can lead to discrimination. First, "proxy" or "second-hand" discrimination describes when the decision-maker differentially weights the relevant information depending on the gender of the candidate, e.g., for a female candidate, language is weighted more heavily. If the intermediaries weigh the attributes differently for candidates that match the gender of the employer compared to candidates that have a different gender than the employer; or if intermediaries differently weigh the attributes for males and females, these differential weighting could partially explain the expected same-gender bias. In Appendix J, we estimate a decision model that allows for differential weights depending on whether the candidates are female or male, or match the gender of the employer (or intermediary). The results show that the weights are not significantly different across groups, and there are no large differences in the female and matching gender biases compared to the baseline model from section 5.1. Therefore, we conclude that proxy discrimination cannot explain our results.

Since participants need to think how another person will make their choices, the cognitive process can be quite complex. Moreover, the cognitive process can also become more difficult when intermediaries hire for employers of the opposite gender, since they might not feel as close to them. We test this possibility by examining the participants' response times and confidence levels in their choices. Appendix section L shows an analysis of the drivers of the participants' confidence and response times (RT). We find no differences in neither confidence or RT when participants hire for someone with the opposite gender compared to their own gender. Moreover, we find that the confidence levels increase and RT decreases for choices where the chosen candidate had higher attributes of *GMA*, *Conscientiousness* and *Interview*. Interestingly, participants report higher confidence levels when they choose a candidate matching the gender of the employer, but do not exhibit shorter RT. These results suggest that even though participants believe the employer will be biased, the decision is not as straightforward as it would be with the other attributes.

7 Conclusions

In this paper, we analyze the role of intermediaries in gender discrimination in hiring. Based on the results of two novel experiments, we find that intermediaries expect the employers (e.g., hiring managers) to discriminate against candidates of the employer's opposite gender. Moreover, when assessing which candidates to forward to the employers, intermediaries will also discriminate against those candidates. We also find that the extent to which the intermediaries discriminate on behalf of the employers is highly heterogeneous, suggesting that these results can be even stronger on the extensive margin.

Although we do not find strong gender discrimination against women, our results uncover a novel source for the gender gap in the hiring market. Since women face difficulties reaching leadership positions, intermediaries in the hiring processes will mostly have to hire for male managers or predominantly male boards. Thus, they will discriminate against female candidates in expectation of same-gender favoritism from the managers. Therefore, our study suggests a new mechanism in which promoting gender-balanced leadership can enhance gender equality.

When we consider what our employers chose, we found that there is no discrimination coming from them. Evidence shows that this type of feedback is very unlikely to reach the intermediaries, thus continuing the expectation of same-gender favoritism. Therefore, companies that are fighting against discrimination need to not only eradicate the discriminatory agents within their workforce, but also need to make sure that the vision of the company has changed as well. Furthermore, if the companies work on developing better feedback structures in their hiring processes, phantom effects of discrimination might be stopped.

Bartlett (2009) describes the current discrimination as practices as having become "more nuanced and pervasive". The mechanisms studied in our paper explain one channel for this phenomenon, namely the expected same-gender favoritism. The identification of this neglected bias allows for future measures to better handle and control the dynamics of the hiring process, ultimately to tackle discrimination.

References

Arrow, K. (1973). The theory of Discrimination. In *Discrimination in Labor Markets*, pages 3–33. Princeton University Press, Princeton, NJ. tex.ids= arrow1971theorya.

Asanov, I. and Mavlikeeva, M. (2023). Can group identity explain the gender gap in the recruitment process? *Industrial Relations Journal*, 54(1):95–113. tex.ids= asanov2023cana, asanov2023canb _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/irj.12392.

Bao, Z. and Huang, D. (2023). Gender-specific favoritism in science. *Journal of Economic Behavior & Organization*.

Barratt, B. (2019). New Survey Begs The Question: Is Your HR Department Sexist? Section: ForbesWomen.

Bartlett, K. T. (2009). Making good on good intentions: The critical role of motivation in reducing implicit workplace discrimination. *Virginia Law Review*, pages 1893–1972. Publisher: JSTOR.

Bartling, B. and Fischbacher, U. (2012). Shifting the Blame: On Delegation and Responsibility. *The Review of Economic Studies*, 79(1):67–87.

Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–75. tex.ids= bartos2016attentiona publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Becker, G. S. (2010). *The economics of discrimination.* University of Chicago press. tex.ids= becker2010economicsa.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. *Handbook of economic field experiments*, 1:309–393. tex.ids= bertrand2017fielda publisher: Elsevier.

Beugnot, J. and Peterlé, E. (2020). Gender bias in job referrals: An experimental test. *Journal of Economic Psychology*, 76:102209.

Bohren, J. A., Imas, A., and Rosenberg, M. (2019). The Dynamics of Discrimination: Theory and Evidence. *American Economic Review*, 109(10):3395–3436. tex.ids= bohren2019dynamicsa publisher: American Economic Association.

Booth, A. L., Francesconi, M., and Frank, J. (1998). *Glass Ceilings or Sticky Floors?* Citeseer.

Booth, A. L., Francesconi, M., and Frank, J. (2003). A sticky floors model of promotion, pay, and gender. *European Economic Review*, 47(2):295–322. Publisher: Elsevier.

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

Brock, J. M. and De Haas, R. (2023). Discriminatory Lending: Evidence from Bankers in the Lab. *American Economic Journal: Applied Economics*, 15(2):31–68.

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203.

Casoria, F., Reuben, E., and Rott, C. (2022). The Effect of Group Identity on Hiring Decisions with Incomplete Information. *Management Science*, 68(8):6336–6345.

Chan, J. and Wang, J. (2018). Hiring Preferences in Online Labor Markets: Evidence of a Female Hiring Bias. *Management Science*, 64(7):2973–2994.

Coffman, K. B., Exley, C. L., and Niederle, M. (2021). The role of beliefs in driving gender discrimination. *Management Science*. tex.ids= coffman2021rolea publisher: INFORMS.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., and Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7(1):25–47.

Drugov, M., Hamman, J., and Serra, D. (2014). Intermediaries in corruption: an experiment. *Experimental Economics*, 17(1):78–99.

Eurostat (2019). The European economy since the start of the millennium - Large enterprises generate one third of employment.

Feld, J., Ip, E., Leibbrandt, A., and Vecci, J. (2022). Identifying and overcoming gender barriers in tech: A field experiment on inaccurate statistical discrimination. Publisher: CESifo Working Paper.

Fenell, A. (2024). Job interview statistics 2024 | UK & Global.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, 44(3):329. Publisher: American Psychological Association.

Focus (2021). Global Gender Equality Survey. Technical report.

Friedmann, E. and Efrat-Treister, D. (2023). Gender Bias in Stem Hiring: Implicit In-Group Gender Favoritism Among Men Managers. *Gender & Society*, 37(1):32–64. Publisher: SAGE Publications Inc.

Glick, P. (1991). Trait-based and sex-based discrimination in occupational prestige, occupational salary, and hiring. *Sex Roles*, 25(5):351–378.

Groot, W. and Van Den Brink, H. M. (1996). Glass ceilings or dead ends: Job promotion of men and women compared. *Economics Letters*, 53(2):221–226. tex.ids= groot1996glassa publisher: Elsevier.

Hagmann, D., Sajons, G., and Tinsley, C. (2022). Statistical Discrimination Against Minority Groups.

Hamman, J. R., Loewenstein, G., and Weber, R. A. (2010). Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship. *American Economic Review*, 100(4):1826–1846.

Hirmas, A. and Engelmann, J. (2024). Learning the value of Eco-Labels: The role of information in sustainable decisions. *Available at SSRN 4788773*.

Hirmas, A., Engelmann, J. B., and Weele, J. v. d. (2024). Individual and contextual effects of attention in risky choice. *Experimental Economics*.

Kausel, E. E., Culbertson, S. S., and Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, 137:27–44.

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

Makles, A. (2012). Stata Tip 110: How to Get the Optimal K-Means Cluster Solution. *The Stata Journal: Promoting communications on statistics and Stata*, 12(2):347–351.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

McKinsey (2024). America's small businesses: Time to think big. Technical report.

Mengel, F., Sauermann, J., and Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European economic association*, 17(2):535–566. tex.ids= mengel2019gendera publisher: Oxford University Press.

Neumark, D., Bank, R. J., and Van Nort, K. D. (1996). Sex discrimination in restaurant hiring: An audit study. *The Quarterly journal of economics*, 111(3):915–941. Publisher: MIT Press.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709. Publisher: JSTOR.

Orquin, J. L. and Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, 144(1):190–206.

Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3):563–602.

Rahal, R.-M. and Fiedler, S. (2019). Understanding cognitive and affective mechanisms in social psychology through eye-tracking. *Journal of Experimental Social Psychology*, 85:103842. tex.ids: rahalUnderstanding2019, rahalUnderstanding2019a.

Rice, L. and Barth, J. M. (2016). Hiring Decisions: The Effect of Evaluator Gender and Gender Stereotype Characteristics on the Evaluation of Job Applicants. *Gender Issues*, 33(1):1–21.

Rosette, A. S. and Tost, L. P. (2010). Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of applied psychology*, 95(2):221. Publisher: American Psychological Association.

Rubinstein, A. and Salant, Y. (2016). "Isn't everyone like me?": On the presence of self-similarity in strategic interactions. *Judgment and Decision Making*, 11(2):168–173. Publisher: Cambridge University Press.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, page 26.

Smith, S. M. and Krajbich, I. (2023). Predictions and choices for others: Some insights into how and why they differ. *Journal of Experimental Psychology: General*, 152(2):528. Publisher: American Psychological Association.

Stepanova, A., Weaver, A., Lahey, J., Alexander, G., and Hammond, T. (2022). Hiring CS Graduates: What We Learned from Employers. *ACM Transactions on Computing Education*, 22(1):1–20.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., and Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, 36(7):778–793. Place: US Publisher: American Psychological Association.

Willemsen, M. C. and Johnson, E. J. (2011). Visiting the decision factory: Observing cognition with MouselabWEB and other information acquisition methods. *A handbook of process tracing methods for decision research*, pages 21–42. Publisher: Psychology Press New York, NY, US.

World Economic Forum (2023). Global Gender Gap Report 2023. tex.ids= worldeconomicforum2023globala.

Ďuriník, M., Morita, H., Servátka, M., and Zhang, L. (2023). Promotions and Group Identity. *Available at SSRN 4656482*.

A Original pre-registered hypotheses

The following table contains all the pre-registered hypotheses of our Beliefs (B), Choice (C) and the Pilot (P) Study[8] The last column shows which hypothesis in the paper reflects the original one.

Table 8: Table with original pre-registered hypotheses

Hyp. in Pre-reg.	Original Hypothesis	Hyp. in Paper
	Pilot	
HP.1	People choose differently when they have to predict a worker's performance and when they predict how others predict a worker performance. This is not only visible in choices, but also shows up in processing data.	Discussion, H1 & H5
	Beliefs Study	
HB.1	The gender of the employers will affect the believed importance of the different applicants' characteristics in their evaluations.	H2
HB.2	The math proficiency of the employers will affect the believed importance of the different applicants' characteristics in their evaluations.	H4
	Choice Study	
HC.1	Female candidates are less likely to be chosen than male candidates.	H1
HC.2	Employers are expected to choose more candidates of their same gender.	H2
HC.3	The proportion of chosen candidates with the same gender of the employer is higher for male compared to female recruiters.	H3
HC.4	Employers with high math skills (compared to low) are expected to put a higher weight on 'General Mental Ability' on their decision.	H4
HC.5	Recruiters that consistently spend more time attending an attribute, weight it more in their decision process.	H6
HC.6	The average time spent looking at the different attributes depends on the employer's characteristics (gender, math level and age).	H5
HC.7	Given the characteristics of a pair of candidates (*GMA*, *Conscientiousness* and *Interview* scores), the proportion of chosen candidates depends on the gender composition (Male vs Female, Female vs Male, Same gender).	H2 & Discussion
HC.8	Given the characteristics of a pair of candidates (*GMA*, *Conscientiousness* and *Interview* scores), differences in the proportion of chosen candidates depend not only on the gender composition, but also on the gender of the employer.	H3 & Discussion

[8]The pre-registrations are available here.

B Pairs of participants for Study 2

In Study 2, participants observe two candidates at the same time and need to choose one of them to send to an employer that they have been assigned to. Due to the high dimensionality of the candidates' data, it is not possible to present all possible combinations to participants. Therefore we needed to determine which combinations to show to them. To do so, we estimated the predicted performance for each candidate based on all employers from our sample (described in section 3.2). With this average predicted performance, we divided each candidate into three groups with equal size: low, medium and high performance.

Both candidates in each pair came from the same performance group. Also, the candidate they would be compared to (by the employer) came from the same group. This ensures that the comparisons are not trivial, since the average predicted performance is relatively similar. Within each performance group, we constructed pairs that fitted the following criteria:

- **Difficulty of comparison** (2 types):
  - *Easy:* One candidate is superior in one attribute (*GMA*, *Conscientiousness* or *Interview*), while the other candidate is superior in another. The third attribute is equal for both.
  - *Complex:* One candidate is superior in two attributes, while the other is superior in the third.
- **Salient attribute** (3 types): In both difficulty cases, there is one candidate that is superior in only one attribute. We created trials in which this attribute is either *GMA*, *Conscientiousness* or the *Interview* score.
- **Gender comparison** (3 types): For each pair of candidates, participants observed one trial where the gender of both candidates was the same (randomized whether both are female/male), one trial in which the better candidate was male and one trial in which the better candidate was female. The values of *GMA*, *Conscientiousness* and *Interview* are identical for each gender pair. Therefore, the statistical comparison between the better candidate being male or female is straightforward. If there is no gender discrimination, the proportion of choices of male/female should be equal across the sample.

Based on our sample, we chose pairs that fitted with the criteria above, yielding 54 trials (3 performance levels $\times$ 2 difficulties $\times$ 3 salient attributes $\times$ 3 gender pairs). Additionally, we added

9 control trials in which the candidates' attributes were identical except for one (*GMA*, *Conscientiousness* or *Interview*), and we repeated those trials for the three gender combinations. We created these trials to control for participants that are not truly paying attention to the information, and choosing randomly instead. In our sample, 361 out of the 391 participants had at least two out of the three control trials correct (and only 2 had none of them correct).

The table below shows all the combinations of trials used in our experiment. Each one of these combinations was presented with the three possible gender combinations. Since we are incentivizing the choices based on what the employers did, the random number for each trial was the same as the one presented to the employers. Therefore, across combinations, these numbers might have differed. Finally, the order of the trials and the side on which the superior candidate was presented (left or right) was randomized, ensuring that our findings are not driven by any positional effects.

Table 9: Attribute combinations for Study 2

#		Candidate A			Candidate B		
		GMA	Conscientiousness	Interview	GMA	Conscientiousness	Interview
C1	Control	5	1	3	5	5	3
C2	Control	4	5	2	4	5	5
C3	Control	1	2	2	2	2	2
1		1	1	4	1	4	3
2		1	1	4	2	3	2
3		1	3	3	2	3	2
4		1	4	5	2	1	3
5		1	5	2	2	2	2
6		1	5	2	2	1	3
7		2	3	3	3	3	2
8		2	3	3	3	1	5
9		2	4	3	3	5	2
10		3	1	5	3	5	3
11		3	4	4	4	1	3
12		3	5	3	4	1	3
13		4	2	3	5	1	4
14		4	2	3	5	3	2
15		4	3	5	4	5	4
16		4	3	5	5	2	4
17		4	5	4	5	1	4
18		4	5	5	5	5	3

C Pilot Study design

This experiment was designed to study whether participants expect others to hold different beliefs than themselves about the performance of the candidates. Participants in this experiment performed two main tasks. Similar to the evaluator-study, participants had to predict the candidates' performance based on the candidate's gender, a personality- and an aptitude tests, and a random number. In the second task, participants needed to predict the predictions of another evaluator. Thus, participants in this study acted both as evaluators, but also had to indicate their beliefs about how others would choose (similar to the intermediary studies)/

This study was done at the CREED Communications Lab at the University of Amsterdam. 35 students (18 females, avg. age 21.5) participated in this experiment. Since this experiment included the usage of an eye-tracker, participants were to bring their contact lenses in case of having an eye prescription, and that anyone with a prescription above ± 3.0 would not be able to take part in the experiment.

Figure 8 shows the trial sequence for both tasks. In each trial, participants observed first the candidate's characteristics, and then they had to predict their performance. For the attributes GMA, Conscientiousness and the interview score, we separated candidates in below or above the median. The random number was defined as odd or even[9]. Therefore, the participants observed 5 binary variables, including the gender of the candidates. For each task, participants observed all the possible attribute combinations (32 trials). The order of the trials was randomized.

[9]This was the first experiment chronologically. Initially, we opted for dividing the attributes into two categories instead of five. Our results suggested that more variation in the attributes can yield better results.

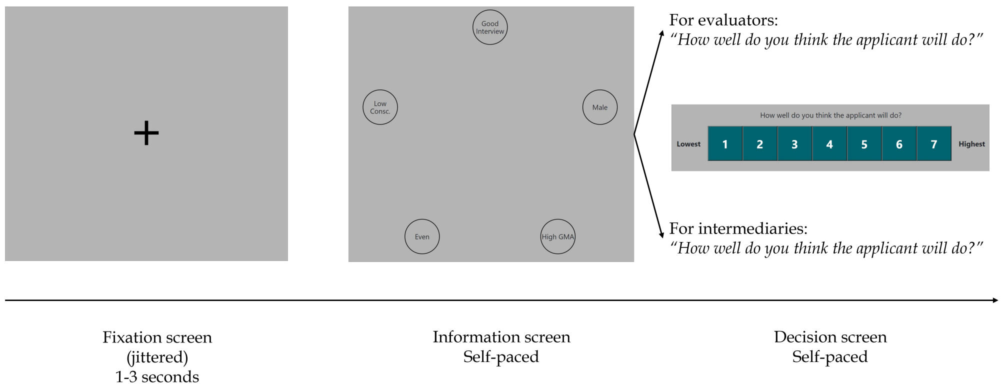

Figure 8: Trial example of Experiment 1

Participants begin the trial by focusing on a fixation cross located in the middle of the screen (for a jittered duration of one to three seconds). Then, we present the information of the candidate. All the attributes are binary: High/Low GMA and Conscientiousness, Good/Bad Interview, Even or Odd number, and Female/Male. When participants are ready, they can press the spacebar on their keyboard to proceed to the decision screen. They can select the score using the Left and Right arrow keys from their keyboard. When they are ready, they can press the spacebar to continue to the next trial.

We recorded the eye movements of the participants via an Gazepoint GP3 HD 150Hz eye-tracker. To be able to properly disentangle the attention of all areas of interest, the attributes were positioned equidistantly from each other. The order of the attributes was randomized at an individual level (i.e., they always saw the attributes on the same location, but the location varied across participants). Before every trial, participants had to fixate their gaze in the middle of the screen by attending a fixation cross. Then, the attributes of the candidate were displayed. Participants could choose at their own pace to proceed to the decision screen.

In the decision screen, participants had predict to which performance group they believed (or thought others believed) the candidate belonged. We explained to participants that we ordered all candidates according to their performance and separated them in seven groups of the same size (instead of five as in the other two experiments). At the end of the experiment, one trial would be chosen. If the participants chose the correct quintile for the candidate, they would get 3 euros as a bonus. In the second task, if their choice matched the modal choice of 10 previous participants, they would get a bonus.

D Moderating effects of employers' math skills

Table 10: Decision models with the intermediaries' math skills

	(1) Beliefs Study	(2) Beliefs Study	(3) Choice Study	(4) Choice Study
Female Candidate	-0.000	-0.015	0.071	0.050
	(0.063)	(0.091)	(0.039)	(0.061)
Medium Math × Female Candidate		0.040		
		(0.088)		
High Math × Female Candidate		-0.008		0.097
		(0.119)		(0.087)
Matching Gender Employer	0.264***	0.336***	0.430***	0.386***
	(0.025)	(0.053)	(0.037)	(0.057)
Medium Math × Matching Gender Employer		0.069		
		(0.059)		
High Math × Matching Gender Employer		0.050		0.106
		(0.065)		(0.081)
Matching Gender Intermediary		0.281**		-0.129
		(0.095)		(0.076)
Medium Math × Matching Gender Intermediary		-0.203*		
		(0.094)		
High Math × Matching Gender Intermediary		-0.318**		0.072
		(0.121)		(0.111)
Matching Gender Both		-0.081		-0.000
		(0.074)		(0.154)
Medium Math × Matching Gender Both		-0.209*		
		(0.089)		
High Math × Matching Gender Both		-0.203*		-0.092
		(0.093)		(0.243)
GMA	0.876***	0.716***	1.168***	1.037***
	(0.023)	(0.029)	(0.047)	(0.061)
Medium Math × GMA		0.181***		
		(0.026)		
High Math × GMA		0.394***		0.261**
		(0.036)		(0.092)
Conscientiousness	0.723***	0.678***	0.800***	0.789***
	(0.024)	(0.033)	(0.029)	(0.040)
Medium Math × Consc.		0.093**		
		(0.030)		
High Math × Consc.		0.128**		0.025
		(0.041)		(0.058)
Interview	0.560***	0.582***	0.859***	0.866***
	(0.024)	(0.036)	(0.036)	(0.051)
Medium Math × Interview		-0.024		
		(0.032)		
High Math × Interview		0.011		-0.008
		(0.043)		(0.072)
Random Number	0.087***	0.114***	0.163***	0.178***
	(0.018)	(0.028)	(0.012)	(0.018)
Medium Math × Random		-0.027		
		(0.025)		
High Math × Random		-0.046		-0.029
		(0.034)		(0.024)
Dependent variable	Second-order Beliefs	Second-order Beliefs	Choices	Choices
Observations	17280	17280	21114	21060
AIC	41770.109	41227.215	20782.520	20699.606
BIC	41863.197	41568.536	20854.139	20946.215
Random-Intercept	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

E Moderating effects of employers' age

Table 11: Decision models with the intermediaries' math skills

	(1) Study 2	(2) Study 2
Female Candidate	0.102*	0.100
	(0.044)	(0.053)
Age(31-40) × Female Candidate		-0.003
		(0.087)
Matching Gender Employer	0.440***	0.382***
	(0.041)	(0.049)
Age(31-40) × Match Gender Employer		0.121
		(0.081)
Matching Gender Intermediary	-0.097	-0.084
	(0.056)	(0.072)
Match Gender Intermediary		-0.016
		(0.111)
Matching Gender Both	-0.067	-0.148
	(0.119)	(0.141)
Matching Gender Both		0.165
		(0.236)
GMA	1.167***	1.208***
	(0.047)	(0.065)
Age(31-40) × GMA		-0.081
		(0.093)
Conscientiousness	0.800***	0.790***
	(0.029)	(0.041)
Age(31-40) × Consc.		0.022
		(0.058)
Interview	0.859***	0.863***
	(0.036)	(0.052)
Age(31-40) × Interview		-0.006
		(0.071)
Random Number	0.163***	0.173***
	(0.012)	(0.017)
Age(31-40) × Random		-0.020
		(0.024)
Dependent variable	Choices	Choices
Observations	21060	21060
AIC	20731.920	20748.683
BIC	20819.426	20995.292
Random-Intercept	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

F Decision models with attention

Table 12: Decision models with attention

	(1) Baseline	(2) Attention	(3) Control Position	(4) Control Int.	(5) Control Both
Decision (choose B)					
Female	0.102*	0.064	-0.009	0.149	0.069
	(0.044)	(0.040)	(0.087)	(0.142)	(0.155)
Female × $\bar{a}(Gender)$		0.063	0.098	0.072	0.108*
		(0.046)	(0.054)	(0.047)	(0.054)
Match Gender E.	0.440***	0.450***	-0.001	0.467***	0.010
	(0.041)	(0.038)	(0.080)	(0.138)	(0.152)
Match Gender Employer × $\bar{a}(Gender)$		0.399***	0.549***	0.403***	0.552***
		(0.043)	(0.052)	(0.045)	(0.052)
Match Gender I.	-0.097	-0.108*	0.056	-0.058	0.138
	(0.056)	(0.055)	(0.129)	(0.239)	(0.274)
Match Gender Intermediary × $\bar{a}(Gender)$		-0.041	-0.081	-0.045	-0.081
		(0.064)	(0.073)	(0.062)	(0.072)
Match Both.	-0.067	-0.090	-0.248	0.371	0.165
	(0.119)	(0.113)	(0.199)	(0.420)	(0.485)
Match Gender Both × $\bar{a}(Gender)$		-0.092	-0.041	-0.094	-0.053
		(0.115)	(0.130)	(0.114)	(0.129)
GMA	1.167***	1.233***	1.219***	1.320***	1.313***
	(0.047)	(0.050)	(0.103)	(0.172)	(0.196)
GMA × $\bar{a}(GMA)$		0.127***	0.152**	0.135***	0.163***
		(0.039)	(0.049)	(0.037)	(0.048)
Conscientiousness	0.800***	0.850***	0.718***	0.694***	0.580***
	(0.029)	(0.030)	(0.049)	(0.104)	(0.113)
Conscientiousness × $\bar{a}(Consc.)$		0.116***	0.183***	0.107***	0.171***
		(0.025)	(0.030)	(0.024)	(0.030)
Interview	0.859***	0.915***	0.896***	0.891***	0.878***
	(0.036)	(0.038)	(0.079)	(0.144)	(0.158)
Interview × $\bar{a}(Interview)$		0.056	0.101*	0.059*	0.096*
		(0.030)	(0.041)	(0.030)	(0.040)
Random	0.163***	0.181***	0.182***	0.083*	0.079*
	(0.012)	(0.012)	(0.012)	(0.037)	(0.038)
Random N. × $\bar{a}(Random)$		0.159***	0.165***	0.165***	0.171***
		(0.014)	(0.014)	(0.014)	(0.014)
Position Controls	No	No	Yes	No	Yes
Intermediaries' Controls	No	No	No	Yes	Yes
Observations	21060	21060	21060	21060	21060
AIC	20731.920	19660.654	19381.424	19560.038	19302.890
BIC	20819.426	19843.622	19803.046	19941.884	19923.390

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G Correlation between attention and attribute importance

In this section, we analyze the raw decisions of the intermediaries without imposing any parametric model. To do so, we calculated for each participant and each attribute, the average proportion of choices favoring a candidate with a higher value on the attribute (or displaying a specific gender). Then, we contrast these proportion of choices favoring the attribute (y-axis) to the individual-average attention (x-axis) in Figure 9. The figure shows how the proportion of choices favoring an attribute compares with the average attention to that specific attribute.

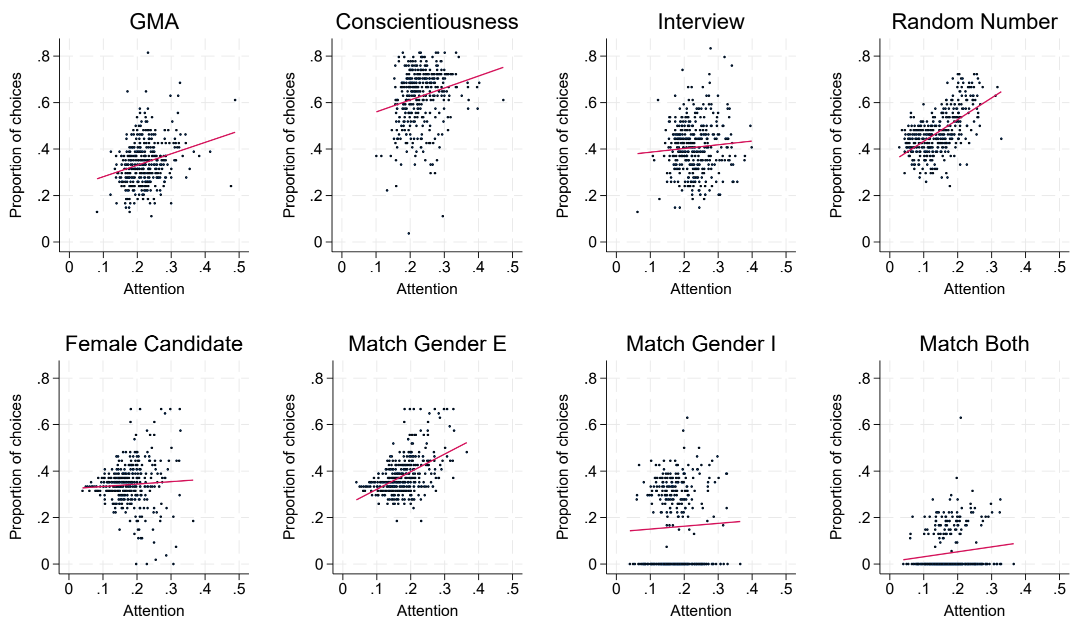

Figure 9: Individual differences in choices and attention

Generally, there is a positive trend on the proportion of choices as the average attention increases. Notably, we find stronger correlations with the proportion of choices matching the gender of the employer (and both the employer and intermediary), and with the random number. Overall, we find a significant Pearson- correlation between choices and attention for *GMA* ($\rho = 0.255$, $p < 0.001$), *Conscientiousness* ($\rho = 0.21$, $p < 0.001$), *Random Number* ($\rho = 0.586$, $p < 0.001$), *matching the gender of the employer* ($\rho = 0.487$, $p < 0.001$) and *Matching both genders* ($\rho = 0.129$, $p < 0.001$). These results support our Hypothesis 5 stating that intermediaries that consistently pay more attention to an attribute are also more likely to choose candidates based on that measure.

H Differences in attention across trials

The Figure below shows the averages in attention for different trials and over time. Panel A shows the attention to the different attributes depending on the attribute composition (i.e. *GMA*, *Conscientiousness*, *Interview* and *Random Number*) on the specific trial. Panel B shows the attention depending on the gender comparison (Equal Gender, Man vs Woman, Woman vs Man). Finally, Panel 3 shows the attention allocated to the different trials depending on the trial number. We use proportion of time spent looking at the attribute as our attention measure. As we can see, there are no strong differences in any of the three panels, thus we conclude that the trial-varying factors do not significantly affect the attention allocated to the different attributes. These results further justify the usage of individual-average attention instead of trial-specific attention.

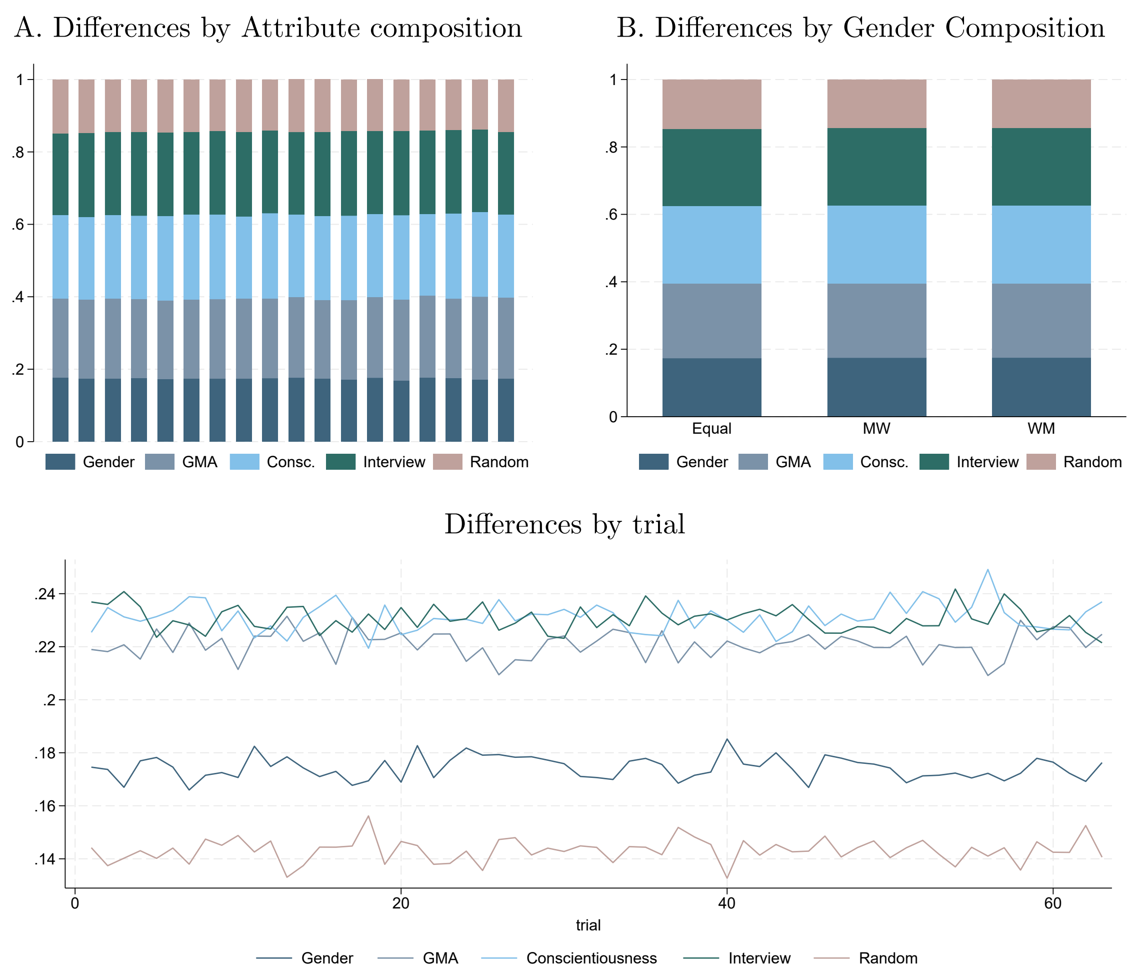

Figure 10: Attention differences across trials

I Attention analysis with other measures

Table 13: Attention model with other measures of attention

	(1) Proportion of Time		(2) ln(DT)		(3) N	
Decision (choose B)						
Female	0.064	(0.040)	0.082	(0.045)	0.101^{*}	(0.045)
Female × $\bar{a}(Gender)$	0.063	(0.046)	-0.018	(0.040)	0.023	(0.036)
Match Gender E.	0.450^{***}	(0.038)	0.486^{***}	(0.044)	0.442^{***}	(0.041)
Match Gender Employer × $\bar{a}(Gender)$	0.399^{***}	(0.043)	0.240^{***}	(0.038)	-0.007	(0.035)
Match Gender I.	-0.108^{*}	(0.055)	-0.078	(0.059)	-0.103	(0.057)
Match Gender Intermediary × $\bar{a}(Gender)$	-0.041	(0.064)	-0.029	(0.063)	-0.072	(0.057)
Match Both.	-0.090	(0.113)	-0.118	(0.121)	-0.018	(0.122)
Match Gender Both × $\bar{a}(Gender)$	-0.092	(0.115)	-0.083	(0.106)	0.164	(0.106)
GMA	1.233^{***}	(0.050)	1.247^{***}	(0.050)	1.180^{***}	(0.047)
GMA × $\bar{a}(GMA)$	0.127^{***}	(0.039)	0.241^{***}	(0.036)	0.094^{*}	(0.037)
Conscientiousness	0.850^{***}	(0.030)	0.853^{***}	(0.031)	0.812^{***}	(0.030)
Conscientiousness × $\bar{a}(Consc.)$	0.116^{***}	(0.025)	0.160^{***}	(0.022)	0.085^{**}	(0.026)
Interview	0.915^{***}	(0.038)	0.918^{***}	(0.039)	0.867^{***}	(0.036)
Interview × $\bar{a}(Interview)$	0.056	(0.030)	0.130^{***}	(0.033)	0.024	(0.031)
Random	0.181^{***}	(0.012)	0.182^{***}	(0.011)	0.167^{***}	(0.012)
Random N. × $\bar{a}(Random)$	0.159^{***}	(0.014)	0.159^{***}	(0.013)	0.060^{***}	(0.015)
Observations	21060		21060		21060	
AIC	19660.654		19766.474		20550.121	
BIC	19843.622		19957.397		20741.044	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

J Second-hand or proxy discrimination

In order to measure whether the participants allocate different decision weights depending on whether the candidates are male or female, or they match the gender of the employer or intermediary, we estimate a model where the attributes *GMA*, *Conscientiousness*, the *Interview* and the *Random number* can have different weights. We use the Oaxaca-Blinder composition (Oaxaca, 1973) to estimate this model. In the table below, we show the baseline model with an added interaction term for the candidates' attributes if the candidate is female (Column 1), matches the gender of the employer (Column 2) or matches the gender of the intermediary (Column 3). As we can see, the estimated differential weights are negligible in scale, and mostly non-significant.

Table 14: Differential decision weights based on gender

	(1) Female vs. Males		(2) Matching E vs Not Matching		(3) Matching I vs Not Matching	
Decision (choose B)						
Female	0.103*	(0.044)	0.102*	(0.044)	0.102*	(0.044)
Match Gender E.	0.432***	(0.041)	0.433***	(0.041)	0.440***	(0.041)
Match Gender I.	-0.136*	(0.063)	-0.130*	(0.063)	0.258	(0.169)
Match Both.	0.003	(0.131)	-0.008	(0.132)	-0.063	(0.119)
GMA	1.168***	(0.047)	1.168***	(0.047)	1.173***	(0.047)
Conscientiousness	0.800***	(0.029)	0.800***	(0.029)	0.803***	(0.031)
Interview	0.859***	(0.036)	0.860***	(0.036)	0.874***	(0.037)
Random	0.163***	(0.012)	0.163***	(0.012)	0.158***	(0.012)
Δ GMA	-0.006	(0.017)	-0.027	(0.015)	-0.035	(0.028)
Δ Conscientiousness	-0.009	(0.013)	0.011	(0.013)	-0.016	(0.038)
Δ Interview	-0.016	(0.018)	0.003	(0.017)	-0.092*	(0.045)
Δ Random Number	0.008	(0.012)	-0.009	(0.012)	0.035	(0.022)
participant.sEmployerGender	-0.036	(0.044)	-0.034	(0.044)	-0.024	(0.043)
_cons	0.030	(0.081)	0.018	(0.081)	-0.045	(0.066)
/						
var(_cons[id])	0.033**	(0.012)	0.033**	(0.012)	0.033**	(0.012)
Observations	21060		21060		21060	
AIC	20735.305		20734.247		20722.820	
BIC	20854.632		20853.574		20842.147	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

K Clustering analysis based on attention

Further, we investigated more complex patterns based on the intermediaries' attention processes. More specifically, we use k-mean clustering to separate the different participants based on the individual-average attention indices for all attributes. We used the methods described in Makles (2012) to determine the optimal amount of cluster. Our data shows that the participants are mainly divided into two clusters. Table 15 shows the average attention allocated to the different information by cluster. The first cluster (N=177) focuses more on the attributes GMA, Conscientiousness and Interview, compared to the second (N=214). The second cluster focuses more on gender and the random number. Due to these differences, we denote the first cluster as "attribute focus" and the second as "gender/random focus". We also compare the demographics of the intermediaries and the treatment conditions across clusters, and we do not find any significant differences in these characteristics across clusters.

Table 15: Individual means by cluster

Cluster	**N**	**p(GMA)**	**p(Consc.)**	**p(Interview)**	**p(Gender)**	**p(Random)**
Attribute Focus	177	0.244	0.257	0.253	0.148	0.098
Gender/Random Focus	214	0.204	0.211	0.212	0.194	0.180

We estimated the baseline model (Section 5.1) moderated by the estimated clusters. Figure 11 shows the estimated decision weights (Panel A) and the gender biases (Panel B) for both clusters. When analyzing the decision weights, we find that the "attribute focus" cluster has significantly larger weights on *GMA* ($p = 0.004$) and *Conscientiousness* ($p < 0.001$), while the "gender/random focus" exhibits a larger weight for the *Random Number* ($p < 0.001$).

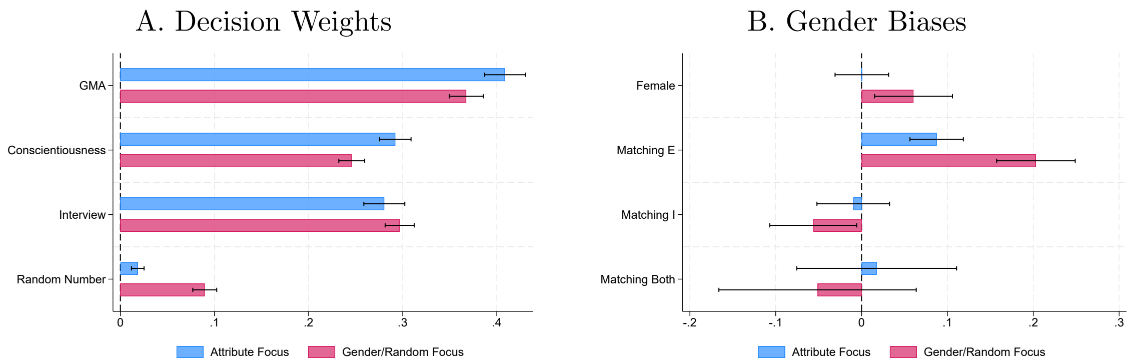

Figure 11: Decision weights and biases

The figures above show the estimated decision weights and biases for Studies 1 (light blue) and 2 (blue), and the employers database (orange). The weights and biases are estimated based on equations (6) using the Delta-method. The parameters come from the estimates from Table 6. Confidence intervals at 95%.

The "gender/random" cluster also exhibits a larger and positive bias in favor female candidates ($p = 0.286$) and the candidates matching the gender of the employer ($p < 0.001$). Although the "gender/random" cluster exhibits a negative bias against candidates matching their own gender, these differences are not significant across groups ($p = 0.1130$).

L Response times and Confidence

The tables below show a linear regression with mixed effects for the Response Times (in natural logarithm) and confidence (standardized at a sample level). In these regressions, we control for whether the intermediary and employer have the same gender to capture any cognitive difficulties on evaluating for someone with a different gender. We also control for the performance level of the candidates. As explained in section B, candidates were separated into three categories based on the predictions of all employers. Finally, we also control for the decisions in which the intermediary chose a candidate that fits a specific gender condition (Female, Matching Gender Employer), or a higher attribute (*GMA*, *Conscientiousness*, *Interview* and *Random Number*). Since RT and confidence have been shown to be closely interrelated, in columns 2 and 4 we control for the respective other variable to capture differential effects that are not captured by differences on the other variable.

Table 16: Regressions on Response Times (RT) and Confidence

	(1) ln(RT)	(2) ln(RT)	(3) Confidence	(4) Confidence
main				
Same Gender (Eval and Int.)	-0.047	-0.045	0.019	0.005
	(0.043)	(0.043)	(0.066)	(0.065)
Low performance	-0.026	-0.005	0.182***	0.175***
	(0.018)	(0.017)	(0.029)	(0.028)
Medium performance	-0.032	-0.016	0.148***	0.138***
	(0.021)	(0.020)	(0.033)	(0.032)
Dec. favoring Female	0.021	0.007	-0.123***	-0.117***
	(0.014)	(0.014)	(0.027)	(0.026)
Dec. favoring matching E	-0.023	-0.002	0.190***	0.183***
	(0.014)	(0.014)	(0.026)	(0.025)
Dec. favoring GMA	-0.111***	-0.079***	0.282***	0.250***
	(0.019)	(0.019)	(0.031)	(0.030)
Dec. favoring Consc.	-0.155***	-0.092***	0.546***	0.502***
	(0.024)	(0.024)	(0.038)	(0.036)
Dec. favoring Interview	-0.176***	-0.122***	0.471***	0.421***
	(0.021)	(0.021)	(0.033)	(0.032)
Dec. favoring Random N.	0.014	0.019	0.041	0.045
	(0.014)	(0.014)	(0.024)	(0.023)
Confidence		-0.115***		
		(0.010)		
ln(RT)				-0.288***
				(0.027)
Observations	4908	4908	4908	4908
AIC	7269.867	7106.526	11805.041	11642.650
BIC	7347.850	7191.008	11883.024	11727.132

Standard errors in parentheses

Confidence ratings are standardized at a sample level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results show no differences on whether the employer and intermediary have the same gender for RT ($p = 0.281$, Column 1) nor Confidence ($p = 0.774$, Column 3). Partipants report higher levels of confidence when comparing medium ($p < 0.001$, Column 3) and high level ($p < 0.001$, Column 3) candidates as compared to low level. Participants also report higher confidence levels and show lower response times when choosing candidates with higher *GMA* ($p < 0.001$, Column 1; $p < 0.001$, Column 3), *Conscientiousness* ($p < 0.001$, Column 1; $p < 0.001$, Column 3), and *Interview* ($p < 0.001$, Column 1; $p < 0.001$, Column 3). Finally, we find that participants do not show lower response times depending on the gender of the candidates, but report lower confidence when choosing *female candidates* ($p < 0.001$, Column 3) and higher confidence when choosing candidates that *match the gender of the employer* ($p < 0.001$, Column 3).