Hiring Algorithm, Statistical Discrimination and Willingness to Invest in Self*

Yung-Shiang Jasmine Yang†

Abstract

We study how individuals respond to algorithmic hiring environments featuring statistical discrimination and uncertain returns to self-investment. In a pre-registered online experiment (N = 553), participants are randomly assigned to an advantaged or disadvantaged identity group and choose how much to invest to improve their chances of being hired by an algorithm with group-based prior beliefs. We find that participants in the disadvantaged group invest 17.7% more than their advantaged counterparts. We also test three interventions to encourage investment: two "Role Model" treatments showing in-group peers who either succeeded or failed after investing, and a "No Risk" treatment that eliminates the financial penalty of investing if not hired. While the role model interventions show limited effects, the "No Risk" treatment significantly increases investment among disadvantaged participants by 24%. Our findings suggest that reducing financial risk, rather than emphasizing representation, may be more effective in motivating self-investment under algorithmic bias.

JEL classification: D91, J71, Z13
Keywords: algorithmic bias, statistical discrimination, human capital, role models, income-contingent loans, online experiment

*I thank Emmanuelle Auriol, Thomas Buser, Haochi Chen, Vessela Daskalova, Paolo Falco, Marco Fongoni, Diogo Geraldes, Emma N. Guirado Forment, Aditi Kharb, Kevin Kloiber, Manuel E. Lago, Lucie Martin, Ivan Pastine, Deirdre Robertson, Mikhail Safronov, Margaret Samahita, Nikita Sangwan, Kun Wu, Bin Zhang, participants at the 2025 European Doctoral Group in Economics (EDGE) Conference, the University College Dublin School of Economics Research & Reading Group Seminar and the 2026 Irish Postgraduate and Early Career Economics (IPECE) Workshop for detailed feedback, helpful suggestions and stimulating conversations. All study procedures of the experiment described were approved by the ethics committee at University College Dublin (081-HS-LR-25-Yang-Samahita). The experiment and hypotheses were pre-registered on AsPredicted under ID #243245 [https://aspredicted.org/f8jk-5xns.pdf]. Funding from University College Dublin School of Economics is gratefully acknowledged.

†School of Economics, University College Dublin, Belfield, Dublin 4, Ireland; E-mail: yung-shiang.yang@ucdconnect.ie

1 Introduction

Algorithms now play an increasingly prominent role in high-stakes decisions, especially in hiring.[1] All of the top 20 Fortune 500 companies rely on data-driven recruitment tools to assist in screening and evaluating candidates (Ajunwa & Greene 2019, Bîgu & Cernea 2019). Employers such as Cisco, Target, and Ikea use algorithms trained on past applications to predict hiring success (Bogen & Rieke 2018, Zhang & Yencha 2022). While these technologies promise increased efficiency, consistency, and reduced taste-based discrimination (Houser 2019, Van Esch et al. 2019), they also raise growing concerns over fairness. In particular, researchers highlight how hiring algorithms may replicate or amplify historical inequalities by relying on group-level priors, a phenomenon known as statistical discrimination (Arrow et al. 1973, Phelps 1972, Barocas & Selbst 2016, Binns 2018).

Statistical discrimination, originally theorized as a rational response to informational constraints, may lead algorithms to encode systematic disadvantages for certain groups, even in the absence of animus. The scalability of algorithmic systems further entrenches these disparities (Kordzadeh & Ghasemaghaei 2022, Awad et al. 2023), as illustrated by Amazon's now-defunct hiring algorithm that penalized resumes containing the word "women" (Hamilton 2018, Vincent 2018). When individuals observe that such systems embed biased expectations, their incentives to invest in themselves may be altered. Theory and empirical evidence suggest such feedback loops can produce self-fulfilling prophecies: where marginalized groups rationally under-invest, confirming the algorithm's priors (Lundberg & Startz 1983, Coate & Loury 1993, Fryer et al. 2005, Fang & Moro 2011, Glover et al. 2017, Patty & Penn 2023). Yet other evidence documents the opposite response towards discrimination: individuals sometimes increase effort in response to perceived bias, a phenomenon known as the "over-efforting" or "twice-as-hard" effect (Caputo 2002, 2007, Evans et al. 2019, Isik et al. 2021, Gutman & Younas 2025).

Building on these strands of studies, this paper investigates how individuals respond to algorithmic hiring environments that embed statistical discrimination and uncertain returns to investment. Specifically, we ask: (1) Do disadvantaged individuals respond differently than advantaged individuals when evaluated by an algorithmic decision-maker with biased group priors? (2) If so, which interventions can effectively encourage self-investment under algorithmic bias?[2]

To answer these questions, we conduct a pre-registered incentivized online experiment (N = 553) that simulates a stylized hiring process with algorithmic screening. Participants are randomly assigned to either a disadvantaged or advantaged identity group and, independently, to a latent productivity type ("top performer" or not) with equal probability (50/50), which is unobservable to both the participant and the algorithm. Participants then choose how much to invest in skill acquisition, which increases their probability of becoming a "top performer," thereby raising the likelihood of achieving higher observable, pre-employment test scores.

[1] We use hiring algorithm as the running example in this paper and in the experiment. The practice of, or the advocacy for, using algorithms as decision-makers is however wider in scope. For example, see Lee & Floridi (2021) on machine learning algorithms making credit risk evaluations and mortgage lending decisions; Kleinberg et al. (2018) on machine learning predictions aiding bail decisions; Cai et al. (2019) on machine learning utilized in image retrieval systems for medical decision making.

[2] It is important to note here the two related but distinct concepts: algorithmic bias and statistical discrimination. Algorithmic bias specifically refers to systematic inequalities embedded within automated systems due to biased training data, algorithmic design choices, or data-processing decisions (Obermeyer et al. 2019, Cowgill et al. 2020). On the other hand, statistical discrimination, a concept rooted in economics, describes scenarios in which decision-makers, human or algorithmic, use group-level information as a proxy for individual characteristics due to uncertainty or information costs (Arrow et al. 1973, Coate & Loury 1993, Fang & Moro 2011). For more detailed distinctions and discussions on the two concepts, see Patty & Penn (2023). This study integrates both perspectives, examining how algorithmic biases shape individual incentives in ways consistent with the theoretical framework of statistical discrimination.

The algorithm does not observe true productivity directly. Instead, it forms hiring decisions based on pre-employment test scores combined with group identity, using group-specific priors about productivity. This is the stage at which algorithmic bias enters: disadvantaged participants face a lower hiring probability as advantaged participants are hired over a wider range of test scores due to more favorable prior beliefs. Participants are informed of this structure and know that their investment might affect their test scores, which in turn could affect hiring outcomes under biased algorithmic evaluation. Conceptually, the design mirrors hiring environments in which applicants belong to groups that are differentially treated in the labor market and invest in credentials or skills to improve observable signals, while algorithmic screeners combine these signals with group identity when making hiring decisions.

We find that disadvantaged participants invest 17.7% more than advantaged participants, consistent with over-efforting rather than under-investment. To shed light on the mechanisms underlying this behavior, we elicit participants' self-reported motivation in a post-experiment questionnaire. Group identity emerges as the most frequently cited factor influencing investment decisions, exceeding considerations such as expected bonus size, investment risk, or information about predecessors' performance. These responses are consistent with compensatory effort or income-targeting reactions to disadvantage (Fongoni 2024), suggesting that participants' awareness of the algorithm's group-based bias, specifically, whether they are assigned to an advantaged or disadvantaged group, plays a central role in shaping investment behavior.

We then test two classes of interventions designed to encourage investment. The first involves information-based encouragement through social comparison: in the "Successful Role Model" and "Unsuccessful Role Model" treatments, participants observe the investment outcomes of five in-group predecessors, motivated by a large literature documenting the effectiveness of role models in reducing bias and shaping aspirations and effort (Dasgupta & Asgari 2004, Beaman et al. 2009, Washington 2008, Beaman et al. 2012). The second intervention targets financial risk reduction: in the "No Risk" treatment, participants only incur the cost of investment if they are hired, mimicking income-contingent repayment schemes.[3] While the role model treatments have limited impact on investment decisions, the "No Risk" treatment increases investment among disadvantaged participants by 24%, suggesting that reducing downside financial risk may be more effective than identity-based messaging in motivating self-investment under algorithmic evaluation.

Our study contributes to several strands of literature. First, we extend the statistical discrimination framework (Arrow et al. 1973, Fang & Moro 2011, Patty & Penn 2023) to an experimental setting involving algorithmic decision-making. In particular, the design of the hiring algorithm in the current study is most closely related to the model in Patty & Penn (2023), which formalizes how algorithms may unintentionally generate discriminatory outcomes through biased priors. Instead of focusing on hiring outcomes or normative assessments of algorithmic fairness, we examine how biased priors shape individuals' costly investment decisions prior to evaluation. Our results lend support to the "over-efforting" or "twice-as-hard" response to disadvantage (Caputo 2002, 2007, Evans et al. 2019, Isik et al. 2021, Gutman & Younas 2025), whereby individuals facing lower expected returns increase effort to compensate. This contrasts with the self-fulfilling prophecy mechanisms, in which anticipated discrimination discourages investment and reinforces initial disparities (Lundberg & Startz 1983, Coate & Loury 1993,

[3]The idea of income-contingent loan schemes can be traced back to Friedman & Kuznets (1945), who proposed a novel approach to financing professional education. Friedman (1955) later expanded the concept to apply more broadly to higher education, arguing that repayments tied to future income could reduce barriers to access while aligning public and private returns on education. The income-contingent model has since been applied in a variety of policy contexts. These include drought relief for farmers (Botterill et al. 2017), where repayments are tied to future income or crop yields; fines for white-collar crimes (Chapman & Denniss 2006), which adjust penalties based on offenders' income; and microfinance schemes for low-income households (Gans & King 2006), where loan repayment schedules are linked to the borrower's capacity to pay.

Fryer et al. 2005, Fang & Moro 2011, Patty & Penn 2023). A key distinction between these responses lies in whether individuals perceive effort as a viable margin of adjustment. In our setting, investment could directly improve observable performance and participants are explicitly informed of this structure, making compensatory effort more likely than disengagement.

Second, this study contributes to a growing literature on how individuals react to algorithmic decision-making, particularly in hiring contexts. Prior research documents skepticism and aversion toward algorithmic evaluators, especially among women, alongside preferences for human decision-makers and concerns about fairness (Dineen et al. 2004, Araujo et al. 2020, Zhang & Yencha 2022). Education and income are positively associated with acceptance of algorithmic screening tools (Zhang & Yencha 2022). Experimental evidence also suggests that discrimination by algorithms may evoke less moral outrage than comparable discrimination by humans (Bigman et al. 2023), and that transparency alone does not necessarily reduce algorithm aversion (Dargnies et al. 2024). While much of this literature focuses on attitudes or stated preferences, fewer studies examine how individuals adjust costly behavior in response to algorithmic evaluation. Notable exceptions study reliance on algorithmic advice and belief updating in judgment tasks (Burton et al. 2020, Chugunova & Sele 2022, Bayer & Renou 2026), as well as labor supply responses at the application stage (Avery et al. 2024). We extend this work by shifting attention to a different behavioral margin: costly, ex-ante investment decisions that are made prior to evaluation and cannot be reversed once an algorithmic assessment takes place. In our incentivized experiment, participants cannot opt out of algorithmic evaluation and must decide how much to invest in skill development in the presence of known algorithmic bias. This distinction between attitudinal responses, reliance on algorithmic advice, and costly investment under non-optional algorithmic evaluation is increasingly important as algorithmic systems expand across hiring, finance, healthcare, and the public sector (Capraro et al. 2024), frequently leaving individuals with little opportunity to avoid algorithmic evaluation.

Third, we contribute to the literature on interventions aimed at mitigating the behavioral consequences of biased evaluation. Existing work highlights the potential of financial instruments such as subsidies, guarantees, and income-contingent loans to counteract disincentives created by discriminatory environments (Chapman 2006, Dianat et al. 2022). Consistent with this perspective, our "No Risk" treatment, which removes downside financial risk by conditioning repayment on hiring, substantially increases investment among disadvantaged participants. In contrast, information-based encouragement through in-group role models has limited impact. This stands in tension with a large literature documenting strong role model effects in human-mediated contexts, including reducing bias (Dasgupta & Asgari 2004, Beaman et al. 2009), raising educational aspirations and attainment (Beaman et al. 2012, Cheryan et al. 2009), shaping political behavior (Washington 2008), and reducing crime (Iyer et al. 2012). A strand of the literature investigates the effect of "negative role models," emphasizing that role model effectiveness depends critically on contextual fit and perceived relevance (Lockwood et al. 2002, ul Amin et al. 2025). Drawing on post-experiment questionnaires, we find that participants primarily cite group identity and financial considerations, rather than predecessor information, as drivers of their investment decisions. Together with the treatment effects, this suggests that algorithmic environments, where outcomes are determined by pre-specified decision rules rather than discretionary judgment, may attenuate the impact of social-comparison interventions while amplifying the effect of concrete payoff-based incentives.

The remainder of the paper proceeds as follows. Section 2 describes the experimental design. Section 3 presents the main results. Section 4 concludes.

2 Experimental Design

The objective of this experiment is twofold. First, we test whether participants randomly assigned to a disadvantaged group invest differently than those assigned to an advantaged group under statistical discrimination from a hiring algorithm. Second, we test the effects of two classes of interventions on investment decisions: information-based encouragement through "Role Model" treatments and financial risk reduction through a "No Risk" treatment. Below we start by describing the information, choices, and payoffs faced by participants and the hiring algorithm in the Control treatment. We then discuss the design of the "Role Model" and "No Risk" treatments, followed by the details of the implementation.

Control treatment: Does random group assignment affect investment?

The Participant's Information. In the experiment, a participant i is randomly assigned to either the Blue or Green identity group, $d_i \in \{b, g\}$. The participant is informed that the hiring algorithm has biased, pre-determined beliefs about each group's productivity as shown in Table 1. Let $y_i \in \{0, 1\}$ denote a participant's productivity, $y_i = 1$ denote that a participant is a top performer, and $y_i = 0$ denote that a participant is an average performer. The algorithm believes that 50% of participants from the Green group are "top performers," while only 20% of participants from the Blue group are believed to be "top performers."[4] Independent of the hiring algorithm's beliefs, the participant understands that whether they are a top performer is in fact randomly assigned: ex ante, each participant has a 50% chance of becoming a top performer and a 50% chance of becoming an average performer.

The participant does not know for certain which type of performer they would be, but understands that there exists a technology of a pre-employment test which serves as an informative but noisy signal to the algorithm. As described in Table 2, a participant's test score x_i takes values in $\{A, B, C\}$ with A being the highest score and C the lowest. The testing technology is such that, conditional on the participant being an average performer, the participant has a 70% chance of scoring the lowest score C and a 30% chance of scoring a neutral score B. Conditional on the participant being a top performer, the participant has a 70% chance of scoring the highest score A and a 30% chance of scoring a neutral score B.

The Participant's Choice. The participant begins with an initial endowment and decides how much to invest, I_i, to enhance their probability of becoming a top performer, subsequently increasing the probability of receiving a higher test score. Every unit of I_i increases the probability that the participant becomes a top performer by 1.5 percentage points. Investment is costly and is deducted from the participant's endowment. If hired by the algorithm, the participant receives positive payoff ω; otherwise, they receive zero.

The Hiring Algorithm's Information and Choice. The hiring algorithm acts as the potential employer and decides whether to hire each participant. It receives net payoff $\pi > 0$ from hiring a top performer, payoff $-\omega < 0$ from hiring an average performer, and zero from not hiring. We exogenously impose that the net benefit of hiring a top performer (π) is 1.5 times the wage (ω) paid to hire a participant. The hiring algorithm's objective is to maximize expected payoffs by only hiring sufficiently likely top performers.

[4]The algorithm's group-based priors are intended to capture statistical discrimination rather than taste-based discrimination. In our setting, the algorithm has no intrinsic preference over groups and seeks only to maximize expected hiring payoffs; group identity enters solely through beliefs about productivity. These beliefs are exogenously imposed to isolate behavioral responses to perceived algorithmic bias, abstracting from how such priors arise in practice (e.g., through biased training data or historical disparities).

Table 1: The hiring algorithm's belief about group productivity rates

	$d_i = b$	$d_i = g$
$y_i = 1$	0.2	0.5
$y_i = 0$	0.8	0.5

Note: $d_i \in \{b, g\}$ indicates whether a participant is in the Blue (b) or Green (g) identity group. y_i denotes a participant's productivity, with $y_i = 1$ denoting that a participant is a top performer, $y_i = 0$ denoting that a participant is an average performer.

Table 2: The group-blind testing technology

	$y_i = 1$	$y_i = 0$
$P(x_i = A \mid y_i)$	0.7	0
$P(x_i = B \mid y_i)$	0.3	0.3
$P(x_i = C \mid y_i)$	0	0.7

Note: x_i denotes a participant's test score which takes values in $\{A, B, C\}$, with A being the highest score and C the lowest. $P(x|y_i)$ denotes the probability that a participant receives a certain test score conditional on their true productivity. For instance, conditional on a participant being a top performer, the probability of them receiving a score of A is 70%, a score of B is 30%, and a score of C is 0%. Conditional on a participant being an average performer, the probability of them receiving a score of A is 0%, a score of B is 30%, and a score of C is 70%.

Formally, the algorithm computes the posterior probability that a participant is a top performer given their group and test score, and hires if this probability exceeds a fixed threshold. The derivation of this threshold is provided in Appendix A.2.

The hiring algorithm's objective implies a simple group-specific hiring policy. Participants with a test score of A are always hired, regardless of group identity, while those with a score of C are never hired. For the intermediate score B, hiring depends on group membership: Green participants are hired, whereas Blue participants are not, reflecting the algorithm's less favorable prior beliefs about the Blue group. Table 3 summarizes the resulting hiring decisions by group and test score.

After the algorithm makes its hiring decision, participants are informed of the outcome and receive their corresponding payoff.

Sequence. The sequence of events in the Control treatment is reiterated below.

1. Each participant i is randomly assigned to either the Blue or Green identity group, $d_i \in \{b, g\}$.

2. The participant learns the hiring algorithm's biased beliefs about group productivity and

Table 3: Algorithmic hiring decisions by group and test score

	$x_i = A$	$x_i = B$	$x_i = C$
$d_i = g$	Hire	Hire	Not hire
$d_i = b$	Hire	Not hire	Not hire

Note: x_i denotes a participant's test score, which takes values in $\{A, B, C\}$, with A being the highest score and C the lowest. Hiring decisions follow from the algorithm's posterior beliefs about productivity conditional on group identity and test score, using a hiring threshold of 40% (see Appendix A.2). Participants with score A are always hired, while those with score C are never hired. For the intermediate score B, Green participants are hired but Blue participants are not, reflecting the algorithm's less favorable prior beliefs about the Blue group.

the testing technology. With their initial endowment, they then decide how much to invest in skill acquisition.

3. After the participant's investment decision, their test score $x_i \in \{A, B, C\}$ is calculated according to Table 2, with $x_i = A$ being the highest score and $x_i = C$ being the lowest.

4. Observing the participant's group d_i and test score x_i, the hiring algorithm decides whether to hire the participant based on the hiring rule as shown in Table 3. The participant learns the hiring decision of the algorithm and their payoff.

Treatment arms: How can participants be encouraged to invest?

The experiment procedure is identical to that in the Control treatment except for the following variations.

"Role Model" treatments: Prior to making their investment decisions, participants are shown the investment decisions and outcomes of five pilot-study participants from their own group (Blue/Green). We randomize participants in this treatment to seeing one of the following two sets of "Role Models": the "Successful Role Models" or the "Unsuccessful Role Models." The "Successful Role Models" are previous participants who invested positive amounts and were hired; the "Unsuccessful Role Models" are previous participants who invested positive amounts and were not hired. Figure 1 shows the screenshots of the 4 potential information screens that participants see. The decisions and outcomes shown in the information screens are based on data from pilot study participants.

"No Risk" treatment: Prior to making the investment decisions, participants are informed that their chosen investment cost is only deducted from their endowment if they are hired, in contrast to in the Control treatment where cost is deducted regardless of hiring outcome. Figure 2 shows the screenshots of the participants' decision screen with the "No Risk" reminder.

Implementation

The study was conducted in August 2025 on Prolific. Participants were recruited to approximate national representativeness of the U.S. population across age, gender and ethnicity.[5] It consists of a practice round, two formal (incentivized) rounds, followed by a demographic questionnaire. Figure 3 depicts the treatment assignment of the experiment. As illustrated in Figure 3, the experiment has both between- and within-subject components. With an estimated effect size of 0.2,[6] and with $\alpha = 0.05$, power $= 0.80$, the power analysis predicts that having 200 observations in each treatment would provide sufficient analytical power.

To achieve the required sample size, under the constraint that we cannot undo an information treatment, we randomly assign each participant to one of the two group identities: Blue or Green, and one of the following four conditions: (1) first round: Control, second round: random assignment to "No Risk", "Successful Role Model" or "Unsuccessful Role Model"; (2) first round: "No Risk", second round: Control; (3) first round: "Successful Role Model", second round: "Unsuccessful Role Model"; (4) first round: "Unsuccessful Role Model", second

[5]Prolific constructs representative samples by matching participant subgroups to U.S. Census Bureau population shares. See `https://researcher-help.prolific.com/en/article/95c345`, accessed 02-17-2026.

[6]The assumed effect size reflects a conservative benchmark, chosen in the absence of closely comparable experimental estimates in algorithmic hiring settings, and is intended to ensure sensitivity to modest treatment effects.

Figure 1: "Successful Role Models" and "Unsuccessful Role Models" information screens

For your reference, before you make your investment decision, here is some information about previous Blue participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Blue	20	A	Yes
Participant B	Blue	20	A	Yes
Participant C	Blue	15	A	Yes
Participant D	Blue	10	A	Yes
Participant E	Blue	5	A	Yes

For your reference, before you make your investment decision, here is some information about previous Green participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Green	20	A	Yes
Participant B	Green	20	B	Yes
Participant C	Green	15	A	Yes
Participant D	Green	12	A	Yes
Participant E	Green	11	A	Yes

For your reference, before you make your investment decision, here is some information about previous Blue participants who have completed this task:

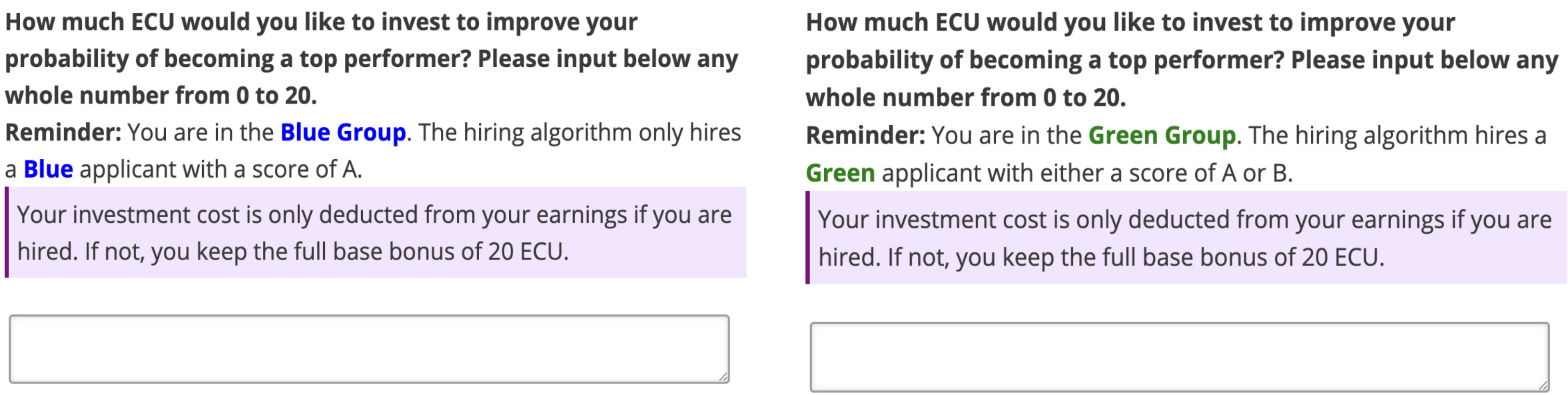
Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Blue	20	B	No
Participant B	Blue	20	C	No
Participant C	Blue	15	B	No
Participant D	Blue	13	C	No
Participant E	Blue	10	C	No

For your reference, before you make your investment decision, here is some information about previous Green participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Green	20	C	No
Participant B	Green	11	C	No
Participant C	Green	10	C	No
Participant D	Green	10	C	No
Participant E	Green	10	C	No

Note: Top left: "Successful Role Models" screen for the Blue group; top right: "Successful Role Models" screen for the Green group; bottom left: "Unsuccessful Role Models" screen for the Blue group; bottom right: "Unsuccessful Role Models" screen for the Green group.

Figure 2: "No Risk" treatment reminders on decision page



Note: left: "No Risk" reminder for the Blue group; right: "No Risk" reminder for the Green group.

round: "Successful Role Model."[7] Across the two decision rounds, we have 275 observations in the Control treatment; 182 in the "No Risk" treatment; 325 in the "Successful Role Model" treatment and 323 in the "Unsuccessful Role Model" treatment. Participants first learn about the relevant information. They then go through a series of attention checks to ensure that they understand the sequence of events, hiring rule and payoff calculations.[8] Afterwards, they undergo a practice round followed by two formal investment rounds.

The survey is concluded with a short demographic questionnaire. We collect demographic variables which include age, gender, education, employment status, income, risk attitude, trust in major technology companies, and whether the participant had interacted with AI or algorithms that served as decision-makers. The full survey is appended in Appendix B.

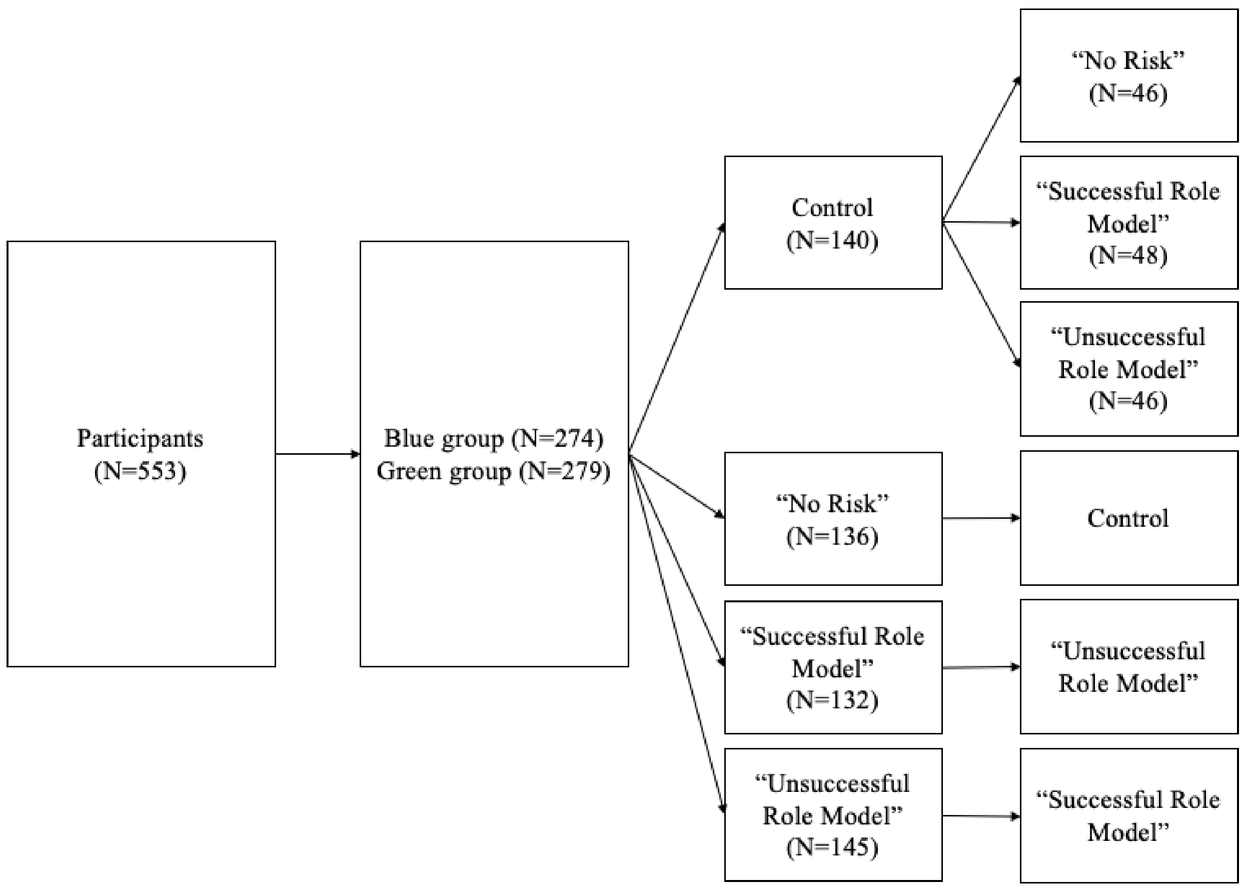


Figure 3: Treatment assignment

The payoff parameters are set as follows. 100 ECUs (Experimental Currency Units) equate 1 USD. At the start of the experiment, each participant is endowed with 20 ECUs to invest in improving their probabilities of becoming a top performer. As described, for every 1 ECU invested, the probability the participant would become a top performer increases by 1.5 percentage points. A participant who is hired by the algorithm receives a wage (ω) of 100 ECUs. A participant's bonus payment from a round is determined by the endowment minus their investment plus their wage if hired. In the "No Risk" treatment, a participant's investment is only deducted from their endowment if they're hired. We randomly select the hiring outcome from one of the two formal rounds to determine the participants' actual bonus payments. The

[7] This design also ensures that a participant would never experience both the "No Risk" and either one of the "Role Model" treatments to avoid confounding the treatment effects.

[8] In the "Role Model" treatments, participants go through additional attention checks answering questions on the previous participants' investment decisions and outcomes before proceeding to making their own investment decisions.

median completion time of the experiment is 9 minutes 21 seconds, each participant receives a participation fee and bonus payment of 2.37 USD on average.

Hypotheses

The purpose of our hypotheses is: (1) to examine whether participants being discriminated against by a hiring algorithm invest differently, and (2) to examine the effects of each treatment on either group of participants. We pre-registered all hypotheses in the AsPredicted registry under ID #243245 [https://aspredicted.org/f8jk-5xns.pdf].

With Hypothesis 1, we first test whether participants in the disadvantaged Blue group invest differently in the presence of algorithmic bias. Under the experimental payoff structure, expected earnings are maximized by investing fully for participants in both groups, regardless of group assignment (see Appendix A.3). Thus, in the absence of behavioral responses to perceived discrimination, standard expected payoff maximization predicts no difference in investment between Blue and Green participants. While the existing literature offers mixed evidence on whether disadvantaged individuals invest more or less than advantaged ones (see, for instance, Glover et al. (2017) and Caputo (2002)), we hypothesize ex ante that there is no difference in investment amounts across groups.

Hypothesis 1 (Group Effect in Control): *In the Control treatment, participants in the disadvantaged Blue group will invest the same as participants in the advantaged Green group.*

With Hypothesis 2, we test the effect of the "Role Model" treatments. Following the findings in the literature on role model effect, we postulate that participants' investments are affected by the "role models" in their group, and that different types of role models could elicit different behavioral response. In the settings of this experiment, "successful" role models refer to previous participants who invested positive amounts and were hired by the algorithm, while the "unsuccessful" role models refer to those who invested positive amounts but were not hired. We hypothesize that being exposed to successful participants encourages investments (Beaman et al. 2012, Cheryan et al. 2009), while being exposed to unsuccessful participants, who invested similarly to the successful ones but were not hired solely due to chance, has a negative impact on participants' willingness to invest (Lockwood et al. 2002, ul Amin et al. 2025).

Hypothesis 2a ("Successful Role Model" Treatment Effect): *Within each group (Blue or Green), participants in the "Successful Role Model" treatment will invest more than those in the Control treatment.*

Hypothesis 2b ("Unsuccessful Role Model" Treatment Effect): *Within each group (Blue or Green), participants in the "Unsuccessful Role Model" treatment will invest less than those in the Control treatment.*

With Hypotheses 3 and 4, we test the effect of the "No Risk" treatment and compare it with that of the Control and "Role Model" treatments. Since payoff variance is lower in the "No Risk" treatment than in the Control treatment, we posit that participants would invest more in the "No Risk" treatment compared to participants in the Control treatment. Similarly, given that the "Role Model" treatments are information treatments that do not alter either payoffs or variances, we hypothesize that participants would also invest more in the "No Risk" treatment compared to in the "Role Model" treatments.

Hypothesis 3 ("No Risk" Treatment Effect): *Within each group (Blue or Green), participants in the "No Risk" treatment will invest more than those in the Control treatment.*

Hypothesis 4 ("No Risk" vs. "Role Model"): *Within each group (Blue or Green), participants in the "No Risk" treatment will invest more than those in either "Role Model" treatment.*

Finally, with Hypothesis 5, we conduct a heterogeneity analysis to examine if there are statistically significant differences in treatment effects across groups. Considering that the "Role Model" and "No Risk" treatments draw on the role model and income-contingent loan literatures, both of which design interventions primarily to benefit disadvantaged groups, we hypothesize that these treatments would have stronger effects on investment decisions among the disadvantaged group.[9]

Hypothesis 5 (Heterogeneity Analysis): *The effects of the "Role Model" and "No Risk" treatments are different across the two groups. Treatment effects will be stronger among the disadvantaged Blue group than the advantaged Green group.*

3 Results

Summary Statistics

Table 4 presents summary statistics. Participants are, on average, 45.8 years old and roughly split between male and female. Approximately 54.5% hold a Bachelor's degree, and 69.3% are employed, with the median income falling within the bracket "Greater than or equal to 50,000 USD and less than 75,000 USD." Participants scored an average of 6.5 on the 0–10 risk attitude scale, and 5.7 on the 0–10 trust scale in response to the question: "In general, how much do you trust major technology companies to make decisions that affect you in a fair and unbiased way?" When asked, "Have you ever applied for a job or other opportunity where you believe your application was screened or reviewed by an AI system or algorithm?" around 35.8% of participants selected the response: "Yes, I have submitted an application where I believe my application was screened by an AI system or algorithm." Out of the participants who selected yes, replying to the follow-up question: "If so, in which contexts have you interacted with the AI system or algorithm? Select all that apply." 96% reported to have had interacted with the AI system or algorithm during job applications or through hiring platforms; 35.4% reported to have had the interaction in the context of credit scoring or during loan applications. We report the summary statistics by identity group (Blue vs. Green), and the summary statistics by treatment (Control, Successful Role Model, Unsuccessful Role Model, and No Risk) respectively in Table A1 and Table A2 in the Appendix. Across both tables, we observe no systematic differences in demographic characteristics, prior experience with AI or algorithm, trust in major tech companies, or risk attitudes.

Figure 4 shows the distribution of investment decisions by group across treatment conditions. Upon first glance, across the four treatments, investments for both groups cluster around 0, 10, and 20 units, with Blue participants more likely to invest the maximum amount of 20. While the distribution of Green participants' investments remains relatively stable, the distribution of Blue participants' investments exhibits sharp spikes at the maximum investment level in the "No Risk" treatment.

[9] Additionally, proof that the disadvantaged Blue group faces higher payoff variances in this experimental setting is collected in Appendix A.3.

Table 4: Summary statistics

Variable	Summary (Mean, SD)
Age	45.79 (15.62)
Risk attitude	6.50 (2.55)
Trust in major tech companies	5.72 (2.40)
Male	49.4%
Employed	69.3%
Prior experience with AI or algorithm	35.8%
Education: Some high school or less	1.1%
Education: High school diploma or GED	13.4%
Education: Some college, but no degree	16.6%
Education: Associates or technical degree	14.3%
Education: Bachelor's degree	35.7%
Education: Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, etc.)	18.8%
Income: Less than 25k	24.2%
Income: 25-50k	25.5%
Income: 50-75k	19.5%
Income: 75-100k	14.6%
Income: 100-125k	4.9%
Income: Greater than or equal to 125k	11.4%
Observations	553

Note: Age is measured in years as a continuous variable. Risk attitude is recorded on a 0–10 scale, with higher scores indicating greater willingness to take risks. Trust in major tech companies is based on responses to the question: "In general, how much do you trust major technology companies to make decisions that affect you in a fair and unbiased way?", also on a 0–10 scale. Male, Employed, and Prior experience with AI or algorithm are binary variables coded as 1 if the participant identify as male, report being employed (full-time or part-time), or have previously submitted an application believed to be screened by an AI system or algorithm.

Effect of Group Advantage on Investment

To test Hypothesis 1, we estimate the following OLS regression using data from the Control treatment:

$$Invest_i = \beta_0 + \beta_1 Blue_i + X_i'\gamma + \varepsilon_i$$

where $Invest_i$ is the outcome variable that denotes the investment amount of participant i; β_0 is the investment level of the Green group; $Blue_i$ is a dummy variable that equals 1 if the participant is in the Blue group, and 0 if the participant is in the Green group. X_i' is a vector of pre-registered controls including round fixed effects, demographics, risk preferences, trust in major tech companies and prior experience with AI or algorithm. Standard errors are clustered at the individual level. Table 5 shows the regression results. In Column 1, we report our main specification controlling only for round fixed effects, in Column 2, we include demographic controls, and in Column 3, risk attitude, prior experience with AI or algorithm and trust in major technology companies are also controlled for.

Across all three specifications, we observe statistically significant, positive effect of being in the Blue group on a participant's investment amount. Our main specification in Column (1) reports that Green participants on average invest 12.4 ECUs out of 20, and that compared to the Green participants, Blue participants invest 2.2 ECUs more. In other words, participants in the disadvantaged Blue group invest around 17.7% more than their counterparts in the advantaged Green group.

This positive effect is also robust under the pre-registered alternative specification, where we exclude participants with duration time outside the 10th and 90th percentile. In another

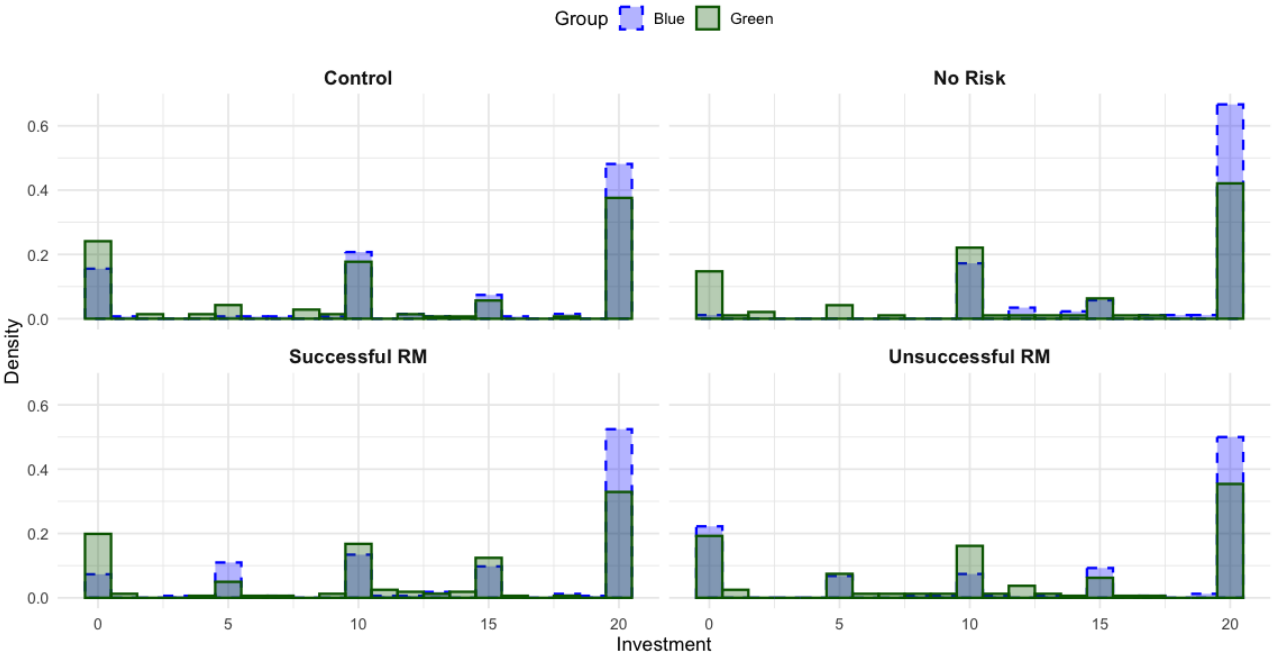


Figure 4: Investment distribution by group

pre-registered alternative specification where we run the analysis using only data from the first decision round, while β_1 is of similar size, it is not statistically significant. This could be due to the analysis being underpowered with only 140 observations. The regression results of the alternative specifications are reported in Table A4 and Table A6 in the Appendix. One additional pre-registered exclusion criterion involves removing participants who failed attention checks. However, due to programming constraints on the survey platform, failed attempts were not recorded once participants eventually provided correct responses. As a result, this exclusion cannot be implemented ex post, and all participants who successfully completed the attention checks are retained in the analysis.

Result 1 (H1 rejected): *In the Control treatment, participants in the disadvantaged Blue group invest more than participants in the advantaged Green group.*

"Role Model" & "No Risk" Treatment Effects

To test Hypotheses 2, 3, 4 & 5, we estimate the following OLS regression:

$$\begin{aligned} Invest_i = \beta_0 + \beta_1 SRM_i + \beta_2 URM_i + \beta_3 NR_i + \beta_4 Blue_i + \beta_5 (SRM_i \times Blue_i) \\ + \beta_6 (URM_i \times Blue_i) + \beta_7 (NR_i \times Blue_i) + X_i' \gamma + \varepsilon_i \end{aligned}$$

where $SRM_i = 1$ if participant i is in the "Successful Role Model" treatment, $URM_i = 1$ if participant i is in the "Unsuccessful Role Model" treatment, $NR_i = 1$ if participant i is in the "No Risk" treatment, $Blue_i = 1$ if participant i is in the Blue group, 0 if otherwise. X_i' is a vector of pre-registered controls including round fixed effects, demographics, risk preferences, trust in major tech companies and prior experience with AI or algorithm. Standard errors are clustered at the individual level. Our coefficients of interest are β_1 and $\beta_1 + \beta_5$ (H2a), β_2 and $\beta_2 + \beta_6$ (H2b), β_3 and $\beta_3 + \beta_7$ (H3), which yield the treatment effects within the Green group and the Blue group respectively and allow us to compare effects across treatments (H4). The coefficients of the interaction terms, $\beta_5, \beta_6, \beta_7$, test heterogeneity in treatment effects across

Table 5: Effect of group (Blue vs. Green) on investment in control treatment

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue Group (vs Green)	2.199**	2.409**	2.184**
	(0.928)	(0.984)	(1.015)
Constant	12.415***	7.961***	8.458**
	(0.820)	(2.668)	(3.788)
Round Control	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	276	276	276
R^2	0.041	0.136	0.235

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Robust standard errors, clustered at the participant level, are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

groups (H5). Table 6 reports the results from our main specification and the linear combination tests.[10]

From Table 6, we first note that both "Role Model" treatments have no statistically significant effect on investment for either group. This finding is consistent with the robustness check reported in Table A5 (Appendix), where we exclude participants whose completion time falls outside the 10th to 90th percentile. However, when we restrict the analysis to only first-round data (Table A7, Appendix), we observe a negative effect of the "Unsuccessful Role Model" treatment on investment for both groups, with participants investing on average 1.9 ECUs less compared to in the Control treatment, while the "Successful Role Model" treatment remains ineffective. This suggests that exposure to information about unsuccessful in-group predecessors may reduce willingness to invest, though the effect appears short-lived.

Result 2 (no support for H2a, H2b): *Within each group (Blue or Green), participants in the "Role Model" treatments do not invest differently than those in the Control treatment in our main specification.*

We next turn our attention to investigating the effect of the "No Risk" treatment. In our main specification, being in the "No Risk" treatment does not affect Green participants' investment decisions significantly. However, for the disadvantaged Blue participants, there is a positive and significant effect. Compared to Blue participants in the Control treatment, being in the "No Risk" treatment increases a Blue participant's investment by approximately 3.5 ECUs (24%). Compared to the Green participants that are also in the "No Risk" treatment, Blue participants in the "No Risk" treatment invest 2.3 ECUs (17%) more.

This positive, heterogeneous effect is also robust under both of our alternative specifications where we exclude participants with duration time outside the 10th and 90th percentile (Table A5) and using data only from the first decision round (Table A7). In the alternative specification where we exclude participants whose completion time falls outside the 10th to 90th percentile (Table A5), we find positive and significant effect of being in the "No Risk" treatment on Green participants' investment amount as well. However, such effect is not detected in the other specification.

Result 3 (partial support for H3): *The "No Risk" treatment has a positive effect on Blue participants' investment amounts, and no effect on Green participants' investment amounts.*

Among Green participants, neither the "Successful Role Model" nor the "Unsuccessful Role Model" treatment significantly affects investment relative to the Control treatment, and the "No Risk" treatment also does not yield a statistically significant change. In contrast, among Blue participants, the interaction term for Blue × No Risk is positive and statistically significant. The linear combination test indicates that Blue participants in the "No Risk" treatment invest significantly more than Blue participants in the Control condition.

To directly compare the "No Risk" treatment with the two "Role Model" treatments within the Blue group, we conduct linear combination tests of the relevant coefficients. These tests show that investment under "No Risk" is significantly higher than under either role model treatment for Blue participants. No such differences are observed within the Green group.

Result 4 (partial support for H4): *Blue participants in the "No Risk" treatment invest*

[10]Table A3 (Appendix) reports treatment effects under all three pre-registered specifications. In Column 1, we report our main specification controlling only for round fixed effects, in Column 2, we include demographic controls, and in Column 3, risk attitude, prior experience with AI or algorithm and trust in major technology companies are also controlled for.

more than those in either "Role Model" treatment. There is no difference in Green participants' investment amount across "No Risk" and "Role Model" treatments.

Result 5 (partial support for H5): *The "Role Model" treatments have no significant effect in either group. The effect of the "No Risk" treatment differs across groups and is only statistically significant among the disadvantaged Blue group.*

Table 6: Treatment effects by group and linear combination tests

	Regression coefficients		Linear combination tests	
	Estimate	Std. Error	Estimate	Std. Error
Main effects				
Blue Group	2.317**	(0.929)		
Successful RM (Green)	0.503	(0.847)	0.503	(0.856)
Unsuccessful RM (Green)	0.152	(0.889)	0.152	(0.856)
No Risk (Green)	1.216	(0.750)	1.216	(0.997)
Interactions				
Blue × Successful RM	0.551	(1.145)	1.055	(0.864)
Blue × Unsuccessful RM	−0.672	(1.232)	−0.520	(0.866)
Blue × No Risk	2.264**	(1.046)	3.480***	(1.022)
Constant	11.744***	(0.678)		
Observations		1,106		
R^2		0.052		

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. The first two columns report OLS regression coefficients from Column (1) of Table A3. The final two columns show linear combination tests comparing each treatment condition to the Control group within Blue and Green participants. Standard errors are clustered at the participant level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Exploratory Analyses

To better understand individual-level heterogeneity in investment decisions, we explore whether demographic characteristics and psychological traits moderate the effects of algorithmic bias and the experimental treatments. We estimate a series of models that include interactions between treatment assignments and participants' age, gender, education, employment status, income, prior experience with algorithms, risk attitudes, and trust in technology companies. Regression results of these analyses are reported in Table A8 to Table A15 in the Appendix.

We find that gender matters: participants identifying as male invest more than the others (Gender coefficient: 1.543, $p < 0.05$), even after controlling for other demographic and psychological variables. Employment status also plays a role. Participants who are currently employed (either full-time or part-time) tend to invest more overall than those who are not (Employment coefficient: 3.740, $p < 0.01$). This positive association is weaker in the "No Risk" treatment (Employment × No Risk: −2.163, $p < 0.1$), though this interaction effect is only marginally significant and should be interpreted cautiously.

Similarly, prior experience with algorithms as decision-makers is positively associated with investment (coefficient: 2.206, $p < 0.01$). Participants with higher education appear to invest marginally more in response to the "Successful Role Model" treatment (Education × Successful RM: 1.954, $p < 0.1$), while participants with higher trust in major technology companies invest marginally less under both the "Successful Role Model" (Trust × Successful RM: 4.026, $p < 0.1$) and "No Risk" (Trust × No Risk: 3.711, $p < 0.1$) treatments.

By contrast, we do not find significant moderating effects for income or general risk attitude, with the exception of a marginally negative interaction between income and the "Successful Role Model" treatment (Income × Successful RM: -0.465, $p < 0.1$). It is also notable that risk attitude neither predicts investment nor significantly interacts with the "No Risk" treatment, despite that the treatment is explicitly designed to mitigate financial downside. This result is consistent with our broader finding that the effectiveness of the "No Risk" intervention does not appear to be driven by baseline risk preferences.

To further probe the mechanisms behind investment decisions, in the post-experiment questionnaire, we ask participants to indicate which factors motivated their choices. Participants can select all that apply out of the following options presented in random order: "My group identity (whether I was assigned to the Green or Blue group)," "Whether I have to pay for the investment cost when I'm not hired," "How much previous participants invested," "Whether previous participants were hired," "The starting amount I had available to invest (20 ECUs)," "The potential bonus I could get if hired (100 ECUs)," "Other (please specify in the textbox below)." Options "How much previous participants invested," and "Whether previous participants were hired" were only displayed to participants who underwent at least one of the "Role Model" treatments. The participants' response separated by group is shown in Figure 5. The most frequently cited factor is group identity (whether the participant is assigned to the Blue or Green group), rather than expected bonus size, investment risks, or information about predecessors' performance or outcome. This suggests that concerns about the algorithm's group-based bias, whether the bias puts a participant in an advantaged or disadvantaged position, may have been the most salient driver of investment decisions.

More specifically, this pattern is consistent with compensatory effort responses to disadvantage, whereby individuals facing lower expected returns increase effort or investment in an attempt to offset unfavorable conditions. In this sense, group identity may serve as a reference point for expected outcomes, motivating higher investment among disadvantaged participants in line with income-targeting or gap-closing mechanisms discussed in the inequality and effort literature (e.g., Fongoni 2024).

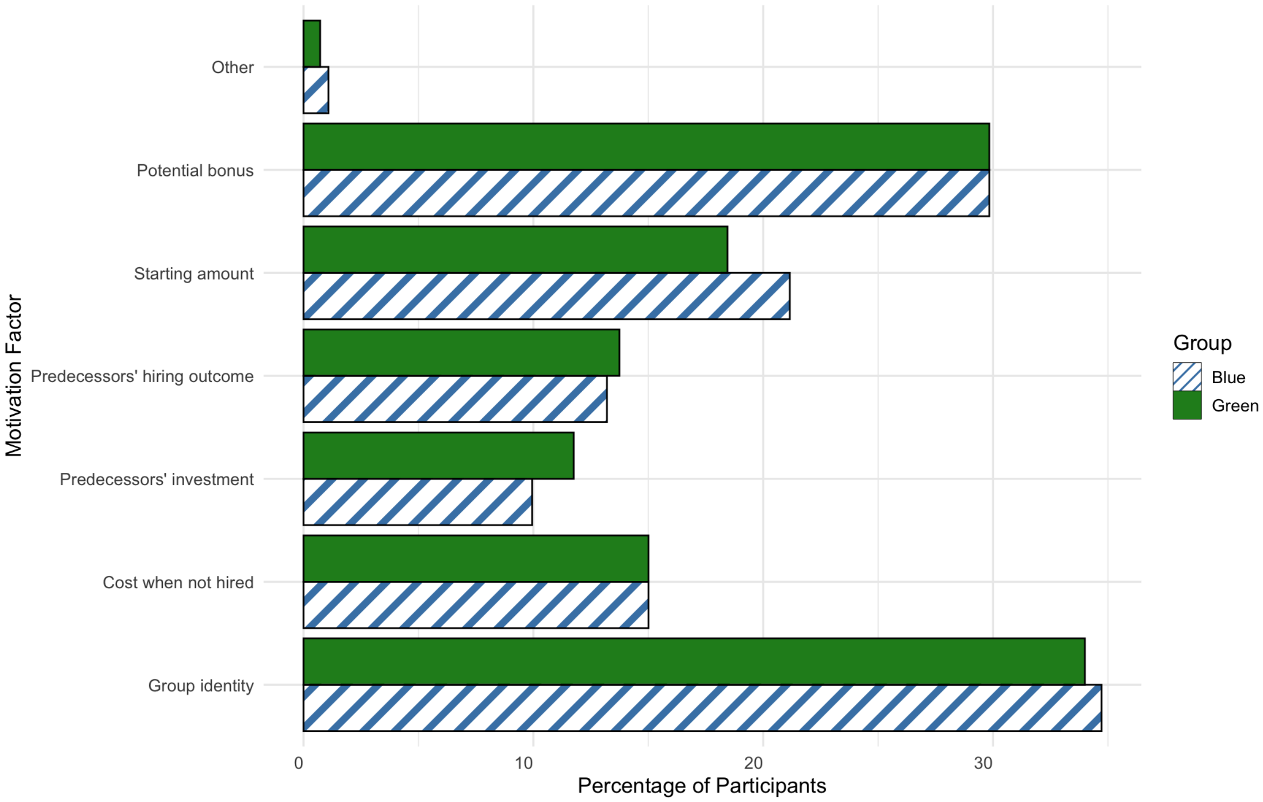
Open-ended responses from participants who selected "Other (please specify in the textbox below)" in the motivation question (Figure 5; $N = 10$) provide additional descriptive insight into participants' reasoning. Given the small size of this subsample, these patterns should be interpreted with caution. We used ChatGPT to group responses into broad thematic categories based on their content, without pre-specifying themes. This process yielded three recurring categories: (1) cost–benefit reasoning, such as "statistics and probability of maximizing my profit" and "risk only 20 cents to potentially gain 1 dollar"; (2) goal-oriented motivations, such as "I wanted to get hired" and "doing whatever it took to increase my likelihood of being hired"; and (3) pessimism and algorithmic aversion, including statements like "no motivation knowing it was a rigged game" and "the AI was designed to not pay out."

4 Conclusion

The results from our experiment reveal that the disadvantaged participants invest more than the advantaged participants, supporting findings from the literature that document the "over-efforting" or "twice-as-hard" effect. We also find that removing the financial risk associated with investing in skill acquisition (via the "No Risk" treatment) leads to higher investment among the disadvantaged participants, while no significant change is observed among the advantaged participants. In contrast, the "Role Model" treatments, designed to influence behavior through social comparison, do not appear to shift investment decisions meaningfully.

Our findings suggest an interpretation in which behavioral responses to discrimination depend not only on disadvantage itself, but also on features of the evaluation environment. In

Figure 5: Self-reported motivation for investment



Note: The exact wording of the question to elicit participants' motivation for investment is: "In the experiment, what factor(s) below motivated your investment choice? Select all that apply." The choices presented in random order that participants can select from are: "My group identity (whether I was assigned to the Green or Blue group)," "Whether I have to pay for the investment cost when I'm not hired," "How much previous participants invested," "Whether previous participants were hired," "The starting amount I had available to invest (20 ECUs)," "The potential bonus I could get if hired (100 ECUs)," "Other (please specify in the textbox below)." Options "How much previous participants invested," and "Whether previous participants were hired" were only displayed to participants who underwent at least one of the "Role Model" treatments.

algorithmic hiring contexts, decision rules are pre-specified and mechanically applied, which limits the scope for discretionary judgment. Relative to human-led selection settings commonly studied in the role model and discrimination literatures, such environments may reduce the effectiveness of social-comparison interventions, while leaving payoff-based incentives salient. Even though the experiment does not include a human-evaluator benchmark, this perspective helps explain the limited impact of the "Role Model" treatments. While social comparison and exposure to successful in-group predecessors have been shown to affect behavior in settings involving human evaluators (see (Washington 2008, Beaman et al. 2009)), such information may be less influential when the evaluator is an algorithm which could be perceived as rigid. In our setting, participants may have interpreted predecessors' outcomes as largely irrelevant for their own prospects, given that the algorithm's hiring rule was fixed and known. As a result, observing successful or unsuccessful in-group predecessors does little to change investment behavior.

By contrast, the strong response to the "No Risk" treatment indicates that interventions which directly modify the payoff structure remain effective under algorithmic evaluation. Removing downside risk increases investment among disadvantaged participants, even when the algorithm's group-based bias remain unchanged. Importantly, this effect does not appear to be driven by baseline risk preferences. This result highlights that individuals facing algorithmic disadvantage are especially responsive to the financial structure of incentives, and that reducing perceived penalties for failure may be more effective than informational nudges in motivating self-investment.

While the experiment offers insights into how algorithmic bias and targeted interventions influence individuals' behavior, it necessarily abstracts from several features of real-world hiring processes. First, the hiring algorithm used in this study applies a fixed decision rule and does not adapt over time. In practice, many algorithmic systems update based on incoming data or operate in hybrid decision environments involving human recruiters (Tambe et al. 2019, Marabelli et al. 2021). A static design was deliberately chosen to maintain experimental control and isolate behavioral responses to perceived algorithmic bias. Allowing the algorithm to adapt would introduce endogeneity between individual investment decisions and evolving hiring thresholds. Future work could extend this framework to dynamic settings in which algorithmic learning and individual behavior co-evolve.

Second, our study abstracts from competitive dynamics among applicants. All participants who meet a predefined qualification threshold may be hired, eliminating strategic considerations about others' behavior. While there are examples of algorithmic screenings being adopted in non-competitive hiring contexts,[11] and that this non-competitive structure allows for clean identification of individual responses to algorithmic bias, future research could examine how these responses change in relative performance or rank-based selection environments. Such extensions would be particularly relevant in labor markets where algorithmic screening is used to allocate scarce opportunities.

Finally, this study focuses on individual behavioral responses rather than institutional reform. While understanding how individuals react to algorithmic disadvantage is informative, it does not address how discriminatory algorithms should be redesigned or regulated. Placing the burden of adjustment on individuals risks reinforcing a "fix the person, not the system" narrative (Bertrand & Duflo 2017, Kline et al. 2022). Nevertheless, individual-level responses are an important input for policy design. Many existing interventions, such as income-contingent loans, training subsidies, or financial guarantees, operate by altering the risk structure faced by individuals. Our findings suggest that such tools may be particularly effective in algorithm-mediated environments, where informational interventions are less salient.

[11] For examples of algorithmic screening outside competitive hiring contexts, see Kleinberg et al. (2018) on bail decisions and Lee & Floridi (2021) on credit risk evaluation.

References

Ajunwa, I. & Greene, D. (2019), Platforms at work: Automated hiring platforms and other new intermediaries in the organization of work, *in* 'Work and labor in the digital age', Vol. 33, Emerald Publishing Limited, pp. 61–91.

Araujo, T., Helberger, N., Kruikemeier, S. & De Vreese, C. H. (2020), 'In ai we trust? perceptions about automated decision-making by artificial intelligence', *AI & society* **35**(3), 611–623.

Arrow, K., Ashenfelter, O. & Rees, A. (1973), 'Discrimination in labor markets', *The Theory of Discrimination* pp. 3–33.

Avery, M., Leibbrandt, A. & Vecci, J. (2024), 'Does artificial intelligence help or hurt gender diversity? evidence from two field experiments on recruitment in tech'.

Awad, E., Balafoutas, L., Chen, L., Ip, E. & Vecci, J. (2023), 'Artificial intelligence and debiasing in hiring: Impact on applicant quality and gender diversity', *Available at SSRN* .

Barocas, S. & Selbst, A. D. (2016), 'Big data's disparate impact', *Calif. L. Rev.* **104**, 671.

Bayer, R.-C. & Renou, L. (2026), 'Interacting with man or machine: When do humans reason better?', *Management Science* **72**(1), 594–608.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R. & Topalova, P. (2009), 'Powerful women: does exposure reduce bias?', *The Quarterly journal of economics* **124**(4), 1497–1540.

Beaman, L., Duflo, E., Pande, R. & Topalova, P. (2012), 'Female leadership raises aspirations and educational attainment for girls: A policy experiment in india', *science* **335**(6068), 582–586.

Bertrand, M. & Duflo, E. (2017), 'Field experiments on discrimination', *Handbook of economic field experiments* **1**, 309–393.

Bigman, Y. E., Wilson, D., Arnestad, M. N., Waytz, A. & Gray, K. (2023), 'Algorithmic discrimination causes less moral outrage than human discrimination.', *Journal of Experimental Psychology: General* **152**(1), 4.

Bîgu, D. & Cernea, M.-V. (2019), Algorithmic bias in current hiring practices: An ethical examination, *in* '13th International Management Conference (IMC) on Management Strategies for High Performance. 13th International Management Conference (IMC) on Management Strategies for High Performance, Bucharest, Romania, October'.

Binns, R. (2018), Fairness in machine learning: Lessons from political philosophy, *in* 'Conference on fairness, accountability and transparency', PMLR, pp. 149–159.

Bogen, M. & Rieke, A. (2018), 'Help wanted: An examination of hiring algorithms, equity, and bias', *Upturn, December* **7**.

Botterill, L. C., Chapman, B. & Kelly, S. (2017), 'Revisiting revenue contingent loans for drought relief: government as risk manager', *Australian Journal of Agricultural and Resource Economics* **61**(3), 367–384.

Burton, J. W., Stein, M.-K. & Jensen, T. B. (2020), 'A systematic review of algorithm aversion in augmented decision making', *Journal of behavioral decision making* **33**(2), 220–239.

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G. S., Stumpe, M. C. et al. (2019), Human-centered tools for coping with imperfect algorithms during medical decision-making, *in* 'Proceedings of the 2019 chi conference on human factors in computing systems', pp. 1–14.

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K. M. et al. (2024), 'The impact of generative artificial intelligence on socioeconomic inequalities and policy making', *PNAS nexus* **3**(6), pgae191.

Caputo, R. K. (2002), 'Discrimination and human capital: A challenge to economic theory and social justice', *J. Soc. & Soc. Welfare* **29**, 105.

Caputo, R. K. (2007), 'Perceived work-related discrimination by women: Implications for social justice and affirmative action', *Journal of Policy Practice* **6**(2), 5–22.

Chapman, B. (2006), 'Income contingent loans for higher education: International reforms', *Handbook of the Economics of Education* **2**, 1435–1503.

Chapman, B. & Denniss, R. (2006), Criminal reparations: using financial incentives and income contingent fines for white-collar crimes, *in* 'Government Managing Risk', Routledge, pp. 176–193.

Cheryan, S., Plaut, V. C., Davies, P. G. & Steele, C. M. (2009), 'Ambient belonging: how stereotypical cues impact gender participation in computer science.', *Journal of personality and social psychology* **97**(6), 1045.

Chugunova, M. & Sele, D. (2022), 'We and it: an interdisciplinary review of the experimental evidence on how humans interact with machines', *Journal of Behavioral and Experimental Economics* **99**, 101897.

Coate, S. & Loury, G. C. (1993), 'Will affirmative-action policies eliminate negative stereotypes?', *The American Economic Review* pp. 1220–1240.

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N. & Chaintreau, A. (2020), Biased programmers? or biased data? a field experiment in operationalizing ai ethics, *in* 'Proceedings of the 21st ACM Conference on Economics and Computation', pp. 679–681.

Dargnies, M.-P., Hakimov, R. & Kübler, D. (2024), 'Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence', *Management Science* .

Dasgupta, N. & Asgari, S. (2004), 'Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping', *Journal of experimental social psychology* **40**(5), 642–658.

Dianat, A., Echenique, F. & Yariv, L. (2022), 'Statistical discrimination and affirmative action in the lab', *Games and Economic Behavior* **132**, 41–58.

Dineen, B. R., Noe, R. A. & Wang, C. (2004), 'Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences', *Human Resource Management: Published in Cooperation with the School of Business Administration, the University of Michigan and in Alliance with the Society of Human Resources Management* **43**(2-3), 127–145.

Evans, D. P., Sales, J. M., Krause, K. H. & Del Rio, C. (2019), 'You have to be twice as good and work twice as hard: a mixed-methods study of perceptions of sexual harassment, assault and women's leadership among female faculty at a research university in the usa', *Global health, epidemiology and genomics* **4**, e6.

Fang, H. & Moro, A. (2011), 'Theories of statistical discrimination and affirmative action: A survey', *Handbook of social economics* **1**, 133–200.

Fongoni, M. (2024), 'Does pay inequality affect worker effort? an assessment of experimental designs and evidence', *Journal of Economic Behavior & Organization* **220**, 697–716.

Friedman, M. (1955), 'The role of government in education'.

Friedman, M. & Kuznets, S. (1945), *Income from independent professional practice*, number frie54-1, National Bureau of Economic Research.

Fryer, R. G., Goeree, J. K. & Holt, C. A. (2005), 'Experience-based discrimination: Classroom games', *The Journal of Economic Education* **36**(2), 160–170.

Gans, J. S. & King, S. P. (2006), Income contingent loans for low-income households, *in* 'Government Managing Risk', Routledge, pp. 216–236.

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

Gutman, L. M. & Younas, F. (2025), 'Understanding the awarding gap through the lived experiences of minority ethnic students: An intersectional approach', *British Educational Research Journal* **51**(2), 990–1008.

Hamilton, I. A. (2018), 'Why it's totally unsurprising that amazon's recruitment ai was biased against women', *Business Insider. Retrieved November* **11**.

Houser, K. A. (2019), 'Can ai solve the diversity problem in the tech industry: Mitigating noise and bias in employment decision-making', *Stan. Tech. L. Rev.* **22**, 290.

Isik, U., Wouters, A., Verdonk, P., Croiset, G. & Kusurkar, R. A. (2021), '"as an ethnic minority, you just have to work twice as hard." experiences and motivation of ethnic minority students in medical education', *Perspectives on medical education* **10**(5), 272–278.

Iyer, L., Mani, A., Mishra, P. & Topalova, P. (2012), 'The power of political voice: women's political representation and crime in india', *American Economic Journal: Applied Economics* **4**(4), 165–193.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. & Mullainathan, S. (2018), 'Human decisions and machine predictions', *The quarterly journal of economics* **133**(1), 237–293.

Kline, P., Rose, E. K. & Walters, C. R. (2022), 'Systemic discrimination among large us employers', *The Quarterly Journal of Economics* **137**(4), 1963–2036.

Kordzadeh, N. & Ghasemaghaei, M. (2022), 'Algorithmic bias: review, synthesis, and future research directions', *European Journal of Information Systems* **31**(3), 388–409.

Lee, M. S. A. & Floridi, L. (2021), 'Algorithmic fairness in mortgage lending: from absolute conditions to relational trade-offs', *Minds and Machines* **31**(1), 165–191.

Lockwood, P., Jordan, C. H. & Kunda, Z. (2002), 'Motivation by positive or negative role models: regulatory focus determines who will best inspire us.', *Journal of personality and social psychology* **83**(4), 854.

Lundberg, S. J. & Startz, R. (1983), 'Private discrimination and social intervention in competitive labor market', *The American economic review* **73**(3), 340–347.

Marabelli, M., Newell, S. & Handunge, V. (2021), 'The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges', *The Journal of Strategic Information Systems* **30**(3), 101683.

Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019), 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science* **366**(6464), 447–453.

Patty, J. W. & Penn, E. M. (2023), 'Algorithmic fairness and statistical discrimination', *Philosophy Compass* **18**(1), e12891.

Phelps, E. S. (1972), 'The statistical theory of racism and sexism', *The american economic review* **62**(4), 659–661.

Tambe, P., Cappelli, P. & Yakubovich, V. (2019), 'Artificial intelligence in human resources management: Challenges and a path forward', *California Management Review* **61**(4), 15–42.

ul Amin, N., Kiani, A., Ahmad, N. & Weng, Q. (2025), 'Motivation by role models: Do they inspire or discourage us? investigating the link between role models and entrepreneurial career intentions', *The International Journal of Management Education* **23**(2), 101090.

Van Esch, P., Black, J. S. & Ferolie, J. (2019), 'Marketing ai recruitment: The next phase in job application and selection', *Computers in Human Behavior* **90**, 215–222.

Vincent, J. (2018), 'Amazon reportedly scraps internal ai recruiting tool that was biased against women', *The Verge* **10**.

Washington, E. L. (2008), 'Female socialization: How daughters affect their legislator fathers' voting on women's issues', *American Economic Review* **98**(1), 311–332.

Zhang, L. & Yencha, C. (2022), 'Examining perceptions towards hiring algorithms', *Technology in Society* **68**, 101848.

A Appendix

A.1 Appendix Tables

Table A1: Summary statistics by group

Variable	Blue	Green	Test differences
Age	46.24	45.34	(0.502)
Risk attitude	6.51	6.48	(0.888)
Trust in major tech companies	5.61	5.83	(0.276)
Male	48.5%	50.2%	(0.701)
Employed	69.3%	69.5%	(0.961)
Prior experience with AI or algorithm	36.1%	35.5%	(0.874)
Education: Some high school or less	1.5%	0.7%	(0.447)
Education: High school diploma or GED	14.6%	12.2%	(0.479)
Education: Some college, but no degree	16.4%	16.8%	(0.985)
Education: Associates or technical degree	12.8%	15.8%	(0.376)
Education: Bachelor's degree	35.4%	36.2%	(0.915)
Education: Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, etc.)	19.3%	18.3%	(0.833)
Income: Less than 25k	25.5%	22.9%	(0.538)
Income: 25-50k	25.5%	25.4%	(1.000)
Income: 50-75k	16.1%	22.9%	(0.053)
Income: 75-100k	13.5%	15.8%	(0.526)
Income: 100-125k	5.5%	4.3%	(0.658)
Income: Greater than or equal to 125k	13.9%	8.6%	(0.068)
Observations	274	279	

Age is measured in years as a continuous variable. Risk attitude is recorded on a 0–10 scale, with higher scores indicating greater willingness to take risks. Trust in major tech companies is based on responses to the question: "In general, how much do you trust major technology companies to make decisions that affect you in a fair and unbiased way?", also on a 0–10 scale. Male, Employed, and Prior experience with AI or algorithm are binary variables coded as 1 if participants identify as male, report being employed (full-time or part-time), or have previously submitted an application believed to be screened by an AI system or algorithm. The column labeled Test differences reports p-values from two-sample t-tests for continuous variables and difference-in-means tests for binary variables, and from chi-squared or Fisher's exact tests for categorical variables (Education and Income).

Table A2: Summary statistics by treatment

Variable	Control	Successful RM	Unsuccessful RM	No Risk	Test differences
Age	45.62	46.03	46.14	45.20	(0.914)
Risk attitude	6.53	6.41	6.48	6.60	(0.872)
Trust in major tech companies	5.80	5.66	5.65	5.89	(0.628)
Male	46.5%	51.4%	52.5%	44%	(0.189)
Employed	68.4%	70.8%	69.9%	68.7%	(0.921)
Prior experience with AI or algorithm	36.7%	35.1%	35.7%	35.2%	(0.977)
Education: Associates or technical degree	14.2%	14.2%	14.9%	12.6%	(0.920)
Education: Bachelor's degree	35.6%	37.5%	34.2%	36.3%	(0.845)
Education: Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, etc.)	18.5%	18.2%	19.3%	19.8%	(0.968)
Education: High school diploma or GED	13.8%	12.6%	14%	13.2%	(0.958)
Education: Some college, but no degree	17.1%	16%	16.5%	17.6%	(0.967)
Education: Some high school or less	0.7%	1.5%	1.2%	0.5%	(0.731)
Income: 100-125k	3.6%	6.5%	5.9%	2.2%	(0.102)
Income: 25-50k	23.6%	27.7%	27.3%	21.4%	(0.324)
Income: 50-75k	23.3%	16.6%	16.5%	24.7%	(0.026)
Income: 75-100k	14.5%	14.2%	14.9%	15.4%	(0.984)
Income: Greater than or equal to 125k	11.3%	10.8%	10.9%	12.6%	(0.925)
Income: Less than 25k	23.6%	24.3%	24.5%	23.6%	(0.992)
Observations	275	325	322	182	

Note: Age is measured in years as a continuous variable. Risk attitude is recorded on a 0–10 scale, with higher scores indicating greater willingness to take risks. Trust in major tech companies is based on responses to the question: "In general, how much do you trust major technology companies to make decisions that affect you in a fair and unbiased way?", also on a 0–10 scale. Male, Employed, and Prior experience with AI or algorithm are binary variables coded as 1 if participants identify as male, report being employed (full-time or part-time), or have previously submitted an application believed to be screened by an AI system or algorithm. The column labeled Test differences reports p-values from one-way ANOVA F-tests of equality across the four treatment groups for continuous and binary variables, and from chi-squared tests (or Fisher's exact tests when expected cell counts are small) for categorical variables (Education and Income).

Table A3: Treatment effects by group

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue Group	2.317**	2.172**	2.014**
	(0.929)	(0.944)	(0.943)
Successful RM (Green)	0.503	0.481	0.286
	(0.847)	(0.854)	(0.861)
Unsuccessful RM (Green)	0.152	0.087	−0.130
	(0.889)	(0.895)	(0.909)
No Risk (Green)	1.216	1.212	1.207
	(0.750)	(0.760)	(0.763)
Blue x Successful RM	0.551	0.735	0.834
	(1.145)	(1.175)	(1.176)
Blue x Unsuccessful RM	−0.672	−0.405	−0.303
	(1.232)	(1.257)	(1.261)
Blue x No Risk	2.264**	2.206**	2.187**
	(1.046)	(1.054)	(1.047)
Constant	11.744***	11.445***	12.897***
	(0.678)	(1.591)	(2.284)
Round Control	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.052	0.074	0.112

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Robust standard errors, clustered at the participant level, are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A4: Effect of group (Blue vs. Green) on investment in control treatment (excluding participants with duration time outside the 10th to 90th percentile)

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue Group (vs Green)	2.311**	2.666**	2.640**
	(1.038)	(1.116)	(1.145)
Constant	12.409***	7.327**	7.831*
	(0.891)	(2.864)	(4.264)
Round Control	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	219	219	219
R^2	0.051	0.197	0.307

Note: This table reports OLS regressions estimating the effect of being assigned to the Blue group (relative to the Green group) on investment decisions in the Control treatment only. The sample excludes participants with completion times below the 10th percentile or above the 90th percentile, following the pre-registered robustness specification. The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Robust standard errors, clustered at the participant level, are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A5: Treatment effects by group (excluding participants with duration time outside the 10th to 90th percentile)

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue Group	2.435**	2.251**	2.121**
	(1.039)	(1.053)	(1.037)
Successful RM (Green)	0.971	0.914	0.799
	(0.948)	(0.950)	(0.954)
Unsuccessful RM (Green)	0.617	0.493	0.388
	(0.970)	(0.973)	(0.983)
No Risk (Green)	1.727**	1.756**	1.643**
	(0.816)	(0.821)	(0.824)
Blue x Successful RM	0.336	0.374	0.560
	(1.286)	(1.307)	(1.305)
Blue x Unsuccessful RM	−1.449	−1.336	−1.149
	(1.381)	(1.399)	(1.387)
Blue x No Risk	2.137*	2.025*	2.192*
	(1.169)	(1.180)	(1.173)
Constant	11.681***	10.172***	11.454***
	(0.742)	(1.790)	(2.440)
Round Control	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	886	886	886
R^2	0.057	0.093	0.142

Note: This table reports OLS regressions estimating treatment effects by group, excluding participants with completion times below the 10th percentile or above the 90th percentile, following the pre-registered robustness specification. The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Robust standard errors, clustered at the participant level, are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A6: Effect of group (Blue vs. Green) on investment in control treatment using only first-round data

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue Group (vs Green)	1.939	1.967	2.167
	(1.259)	(1.384)	(1.413)
Constant	12.554***	8.558***	5.265
	(0.921)	(3.063)	(4.805)
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	140	140	140
R^2	0.017	0.154	0.358

Note: This table reports OLS regressions estimating the effect of group assignment (Blue vs. Green) on investment in the Control treatment, using only first-round observations as specified in the pre-registration. The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). Column (2) adds demographic controls: age (continuous), gender (binary), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Robust standard errors, clustered at the participant level, are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A7: Treatment effects using only first-round data

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue Group	1.939	1.458	1.422
	(1.209)	(1.245)	(1.243)
Successful RM (Green)	1.087	0.739	0.568
	(1.256)	(1.289)	(1.297)
Unsuccessful RM (Green)	−1.878	−2.079*	−2.439*
	(1.213)	(1.248)	(1.264)
No Risk (Green)	−0.475	−0.653	−0.895
	(1.205)	(1.243)	(1.271)
Blue x Successful RM	0.302	1.208	0.987
	(1.733)	(1.800)	(1.806)
Blue x Unsuccessful RM	0.357	0.825	1.213
	(1.693)	(1.757)	(1.765)
Blue x No Risk	3.048*	3.305*	3.261*
	(1.726)	(1.789)	(1.810)
Constant	12.554***	12.491***	12.562***
	(0.885)	(1.895)	(2.591)
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	553	553	553
R^2	0.069	0.099	0.154

Note: This table reports OLS regressions estimating treatment effects by group using only first-round observations, following the pre-registered robustness specification. The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). Column (2) adds demographic controls: age (continuous), gender (binary), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Robust standard errors, clustered at the participant level, are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A8: Treatment effects by age

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.654***	2.651***	2.539***
	(0.547)	(0.555)	(0.545)
Successful RM	0.061**	0.056*	0.062*
	(0.031)	(0.033)	(0.033)
Unsuccessful RM	3.568**	3.418*	3.272*
	(1.796)	(1.815)	(1.815)
No Risk	1.860	1.855	1.681
	(1.842)	(1.856)	(1.837)
Age	3.251**	3.260*	2.995*
	(1.645)	(1.670)	(1.665)
Age × Successful RM	−0.061	−0.056	−0.056
	(0.037)	(0.038)	(0.038)
Age × Unsuccessful RM	−0.045	−0.043	−0.043
	(0.038)	(0.038)	(0.038)
Age × No Risk	−0.020	−0.021	−0.016
	(0.034)	(0.035)	(0.035)
Constant	8.771***	9.798***	11.275***
	(1.543)	(1.905)	(2.498)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.054	0.073	0.111

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: gender (binary), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Age is measured in years as a continuous variable. Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A9: Treatment effects by gender

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.657***	2.628***	2.528***
	(0.548)	(0.555)	(0.546)
Successful RM	−1.349	−1.526	−1.693*
	(0.934)	(0.952)	(0.948)
Unsuccessful RM	−0.117	−0.023	−0.074
	(0.777)	(0.779)	(0.782)
No Risk	−1.233	−1.124	−1.189
	(0.865)	(0.867)	(0.865)
Gender	1.551**	1.568**	1.543**
	(0.682)	(0.691)	(0.686)
Gender × Successful RM	1.859	1.842	1.629
	(1.148)	(1.149)	(1.153)
Gender × Unsuccessful RM	2.133*	2.082*	1.861
	(1.235)	(1.235)	(1.232)
Gender × No Risk	1.626	1.520	1.542
	(1.058)	(1.079)	(1.087)
Constant	12.205***	12.041***	13.883***
	(0.676)	(1.491)	(2.114)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.052	0.074	0.111

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), education level (ordered factor), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Gender is coded as 1 if the participant identifies as male, 0 if the participant identifies as female of non binary/third-gender. Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A10: Treatment effects by education

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.778***	2.716***	2.629***
	(0.548)	(0.552)	(0.544)
Successful RM	2.382	1.772	1.661
	(3.785)	(3.955)	(3.827)
Unsuccessful RM	−2.406	−2.877	−2.953
	(3.490)	(3.613)	(3.473)
No Risk	−0.365	−0.330	−0.248
	(2.515)	(2.592)	(2.497)
Education	1.012	0.676	0.581
	(1.599)	(1.632)	(1.575)
Education × Successful RM	1.899	1.873	1.954*
	(1.210)	(1.190)	(1.168)
Education × Unsuccessful RM	1.000	1.132	0.993
	(1.124)	(1.148)	(1.098)
Education × No Risk	0.693	0.826	0.641
	(1.310)	(1.335)	(1.264)
Constant	11.139***	10.638***	11.972***
	(1.147)	(1.754)	(2.403)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.071	0.087	0.126

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), employment status (binary), and income bracket (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Education is an ordinal variable ranging from 1 (some high school or less) to 6 (graduate or professional degree). Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A11: Treatment effects by income

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.651***	2.646***	2.612***
	(0.549)	(0.546)	(0.545)
Successful RM	0.766	0.751	0.756
	(0.581)	(0.582)	(0.581)
Unsuccessful RM	−0.197	−0.223	−0.228
	(0.618)	(0.620)	(0.619)
No Risk	2.288***	2.294***	2.271***
	(0.522)	(0.520)	(0.519)
Income	0.231	0.225	0.218
	(0.212)	(0.213)	(0.210)
Income × Successful RM	−0.465*	−0.445*	−0.447*
	(0.254)	(0.252)	(0.250)
Income × Unsuccessful RM	−0.328	−0.314	−0.307
	(0.274)	(0.273)	(0.273)
Income × No Risk	0.028	0.030	0.037
	(0.237)	(0.238)	(0.237)
Constant	14.322***	13.319***	11.749***
	(0.941)	(1.634)	(1.978)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.078	0.092	0.134

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education (ordered factor), and employment status (binary). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Income is a participant's gross annual salary, an ordinal variable ranging from 1 (less than 25,000 USD) to 13 (greater than/equal to 300,000 USD). Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A12: Treatment effects by employment status

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.648***	2.564***	2.453***
	(0.549)	(0.558)	(0.547)
Successful RM	−0.169	−0.469	−0.429
	(1.041)	(1.092)	(1.080)
Unsuccessful RM	0.745	0.597	0.438
	(1.107)	(1.110)	(1.109)
No Risk	−0.029	−0.161	−0.283
	(1.144)	(1.147)	(1.147)
Employment	3.721***	3.650***	3.740***
	(0.933)	(0.925)	(0.921)
Employment × Successful RM	0.054	0.368	0.388
	(1.284)	(1.295)	(1.300)
Employment × Unsuccessful RM	−0.224	0.067	0.013
	(1.356)	(1.365)	(1.366)
Employment × No Risk	−2.070*	−1.997*	−2.163*
	(1.137)	(1.133)	(1.127)
Constant	11.705***	12.271***	13.406***
	(0.916)	(1.741)	(2.459)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.052	0.068	0.107

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education (ordered factor), and income (ordered factor). Column (3) further includes individual-level covariates: risk attitude (self-reported on a 0–10 scale), prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Employment status is coded as 1 if the participant reports to be working full-time or part-time, 0 if the participant reports to be unemployed and looking for work, a homemaker or stay-at-home parent, a student or retired. Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A13: Treatment effects by risk attitude

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.587***	2.516***	2.439***
	(0.546)	(0.554)	(0.552)
Successful RM	−4.240	−4.496	−4.535
	(3.358)	(3.155)	(3.346)
Unsuccessful RM	−3.616	−3.864	−3.116
	(2.658)	(2.569)	(2.683)
No Risk	0.385	0.539	1.052
	(2.299)	(2.257)	(2.385)
Risk attitude	−1.720	−1.540	−1.293
	(2.475)	(2.391)	(2.526)
Risk attitude × Successful RM	0.311	0.544	1.080
	(2.220)	(2.165)	(2.361)
Risk attitude × Unsuccessful RM	−2.319	−2.374	−1.624
	(2.021)	(1.967)	(2.140)
Risk attitude × No Risk	−1.835	−2.195	−1.673
	(2.103)	(2.078)	(2.247)
Constant	13.008***	12.246***	12.486***
	(1.819)	(2.342)	(2.776)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.081	0.107	0.126

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education (ordered factor), employment status (binary), and income (ordered factor). Column (3) further includes individual-level covariates: prior experience with algorithmic screening (binary), and trust in major technology companies (0–10 scale). Risk attitude is the participant's response on a 0 to 10 scale to the question: "In general, how willing are you to take risks? (0 = Completely unwilling to take risks; 10 = Very willing to take risks)." Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A14: Treatment effects by trust in major tech companies

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.674***	2.610***	2.546***
	(0.558)	(0.560)	(0.551)
Successful RM	0.182	−0.586	−0.933
	(2.779)	(2.824)	(2.805)
Unsuccessful RM	0.508	−0.214	−0.468
	(2.548)	(2.554)	(2.523)
No Risk	0.747	−0.260	−0.644
	(2.319)	(2.285)	(2.307)
Trust	0.530	−0.070	−0.166
	(2.079)	(2.057)	(2.080)
Trust × Successful RM	−2.996	−3.787*	−4.026*
	(2.199)	(2.140)	(2.180)
Trust × Unsuccessful RM	−0.920	−1.727	−2.401
	(2.314)	(2.280)	(2.337)
Trust × No Risk	−2.365	−3.115	−3.711*
	(2.201)	(2.153)	(2.193)
Constant	12.363***	12.573***	13.184***
	(1.873)	(2.420)	(2.738)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Prior Experience	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.074	0.100	0.124

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education (ordered factor), employment status (binary), and income (ordered factor). Column (3) further includes individual-level covariates: risk attitude (0-10 scale), and prior experience with algorithmic screening (binary). Trust is the participant's response on a 0 to 10 scale to the question: "In general, how much do you trust major technology companies to make decisions that affect you in a fair and unbiased way? (0 = No trust at all; 10 = Complete trust)." Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table A15: Treatment effects by prior experience with AI/algorithms

	Dependent variable:		
	Investment		
	(1)	(2)	(3)
Blue	2.660***	2.631***	2.573***
	(0.549)	(0.554)	(0.547)
Successful RM	−0.866	−0.922	−1.276
	(0.959)	(0.979)	(0.996)
Unsuccessful RM	0.624	0.702	0.522
	(0.720)	(0.729)	(0.716)
No Risk	−0.102	−0.050	−0.249
	(0.788)	(0.799)	(0.780)
Prior experience	2.255***	2.228***	2.206***
	(0.643)	(0.653)	(0.657)
Prior experience × Successful RM	0.374	0.402	0.685
	(1.195)	(1.202)	(1.209)
Prior experience × Unsuccessful RM	−0.273	−0.182	0.080
	(1.260)	(1.265)	(1.263)
Prior experience × No Risk	0.118	0.112	0.095
	(1.108)	(1.114)	(1.114)
Constant	11.875***	11.928***	13.819***
	(0.651)	(1.635)	(2.245)
Round Fixed Effects	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes
Risk Attitude	No	No	Yes
Trust in Tech	No	No	Yes
Observations	1,106	1,106	1,106
R^2	0.051	0.073	0.106

Note: The dependent variable is the number of investment units allocated (ranging from 0 to 20 inclusive). All models include round fixed effects. Column (2) adds demographic controls: age (continuous), gender (binary), education (ordered factor), employment status (binary), and income (ordered factor). Column (3) further includes individual-level covariates: risk attitude (0-10 scale), and trust in major technology companies (0–10 scale). Prior experience is coded as 1 if the participant responds to the question: "Have you ever applied for a job or other opportunity where you believe your application was screened or reviewed by an AI system or algorithm? (For example: resume filtering by applicant tracking systems used by companies and recruitment agencies; or credit checks by systems like FICO Score.)" by selecting the option: "Yes, I have submitted an application where I believe my application was screened by an AI system or algorithm." 0 if otherwise. Robust standard errors clustered at the participant level in parentheses. *p<0.1; **p<0.05; ***p<0.01.

A.2 Algorithmic Hiring Rule Derivation

Let $\pi > 0$ denote the hiring algorithm's net payoff from hiring a top performer, $-\omega < 0$ the payoff from hiring an average performer (or, the wage of hiring any participant), and 0 when not hiring. Observing the participant's group (d_i) and test score (x_i), the hiring algorithm hires the participant if and only if the expected utility of hiring, $EU(hire)$, satisfies:

$$EU(hire) = P(y_i = 1|d_i, x_i) \times \pi + (1 - P(y_i = 1|d_i, x_i)) \times (-\omega) > 0$$

That is, to be hired by the algorithm, $P(y_i = 1|d_i, x_i)$ must satisfy:

$$P(y_i = 1|d_i, x_i) \geq \frac{\omega}{(\pi + \omega)}$$

In the experiment, we exogenously impose that the net benefit of hiring a top performer (π) is 1.5 times the wage (ω) paid to hire a participant. Plugging the numbers in the above function, the hiring algorithm hires a participant when, conditional on the participant's group and test score, the probability that a participant is a top performer is no less than 40%.

When a participant receives a test score of A, with the group-blind testing technology described in Table 2, the hiring algorithm could be 100% certain that this participant is a "top performer" and would hire regardless of the participant's group identity. Similarly for when a participant receives a test score of C, the hiring algorithm could be 100% certain that the participant is an "average performer" and would never hire regardless of the participant's group. When a participant receives a B, however, given that it is an uninformative score (both "top performers" and "average performers" are equally likely to receive a score of B), the hiring algorithm has to refer to its biased group-based priors to make a decision. Recall from Table 1 the hiring algorithm's belief about each group's productivity rate, where it believes that 50% of participants from the Green group are "top performers" and that only 20% of participants from the Blue group are "top performers." For a Green participant with a score of B, the hiring algorithm rely on its priors and believes that the participant has a 50% chance of being a "top performer," and would hire the participant given that the probability is higher than the hiring threshold of 40%. However, for a Blue participant with a score of B, given that the hiring algorithm's prior of Blue participants being "top performers" is 20%, they would not be hired given that the probability is lower than the hiring threshold.

A.3 Marginal Payoff and Variance Derivation

With the testing technology and the hiring threshold described in Section 2 and Appendix A.2, and denote $p_0 \in \{p_g, p_b\}$ as a participant's baseline probability of being hired without investment, a Green participant's baseline probability of being hired (p_g) is 0.65; a Blue participant's baseline probability of being hired (p_b) is 0.35. In other words, a Green participant's baseline probability of scoring either an A (0.35) or a B (0.3), and a Blue participant's baseline probability of scoring an A (0.35). Recall that investing I_i increases a participant's chance of being the top performer by 1.5%, incorporating the endogenous investment decisions, a green participant's probability of being hired (p) can thus be expressed as: $(0.5+0.015I_i)\times 1+(1-0.5-0.015I_i)\times 0.3 = 0.65+0.0105I_i$. For a Blue participant, they are only hired when they have a score of A. Their probability of being hired (p) can therefore be expressed as: $(0.5+0.015I_i)\times 0.7 = 0.35+0.0105I_i$. Subsequently, we can express participant i's probability of being hired as a function of their baseline hiring probability p_0 and their investment (in ECU) I_i as:

$$p = P(hired) = p_0 + 0.0105I_i$$

The expected payoff (in ECU) of investing I_i in the Control treatment (μ_C) can thus be expressed as:

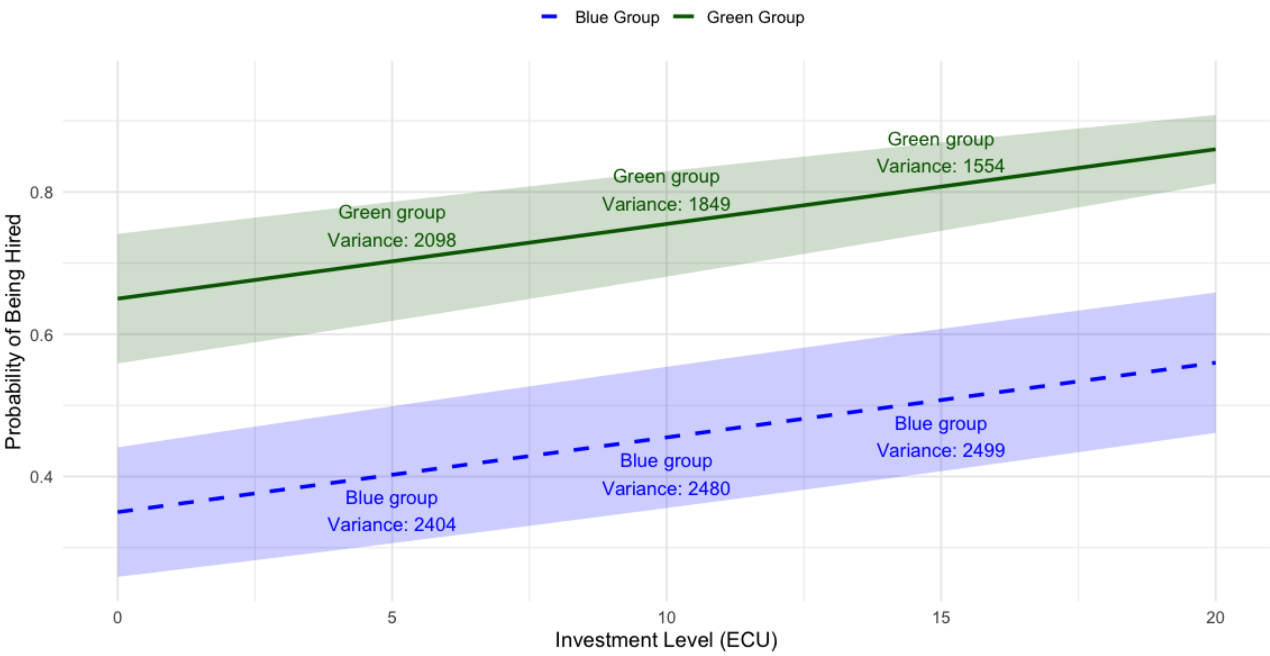
$$\begin{aligned}\mu_C &= p(100-I_i)+(1-p)(-I_i)\\ &= 100p - I_i\end{aligned}$$

which is then straightforward to calculate that the expected payoff for a Green participant is $65+0.05I_i$, and $35+0.05I_i$ for a Blue participant. This investment–qualification relationship is calibrated so that the marginal benefit of each additional ECU invested exceeds its cost. Across both groups, participants' expected payoffs are maximized when they invest the full allowable amount of 20 ECUs.

Despite that the marginal payoff of investing is the same across the two groups, the stricter hiring rule faced by the Blue participants makes their return on investment riskier. Figure A1 visualizes the probability of being hired, the expected payoffs, and the variance of payoffs in the Control treatment. As is made apparent in the figure, Blue participants face higher variance in payoff across all investment amounts compared to Green participants.

This informs our design of the "No Risk" treatment, where participants only pay their chosen investment cost if they are successfully hired, reducing the variance associated with investing in improving one's outcome. Given that participants do not incur investment costs if they're not hired in the "No Risk" treatment, their expected payoff μ_{NR} can be expressed as $p(100-I_i)+(1-p)(0) = p(100-I_i)$. Let VAR_C and VAR_{NR} denote the variance of payoff in the Control and the "No Risk" treatment. We can express VAR_C as $p(100-I_i-\mu_C)^2+(1-p)(-I_i-\mu_C)^2$; VAR_{NR} as $p(100-I_i-\mu_{NR})^2+(1-p)(0-\mu_{NR})^2$. Given that p is no greater than 1, μ_{NR} is strictly greater than μ_C, VAR_C is therefore also strictly higher than VAR_{NR}.

Figure A1: Hiring probability and payoff variance by group in the Control treatment



Note: Investment level in ECU is on the x-axis, probability of being hired is on the y-axis. The green (blue dash) curve and shade are the Green (Blue) participant's hiring probability and variance of payoff.

B Full Survey

Welcome & Consent

Welcome and Study Information

You are invited to take part in a short research study on decision-making, conducted by a researcher from the School of Economics at University College Dublin.

What you'll do:

- You will be given some information about a scenario and then you will be asked to make a number of decisions
- You have the chance to earn an additional bonus depending on your decisions
- Answer a brief demographic questionnaire

Participation details:

- You must be at least **18 years old**
- The study takes about **10 minutes**
- On top of the participation fee, you will earn a bonus between **$0 USD** to **$1.2 USD** based on your decisions

Data & Privacy:

- Your responses are **anonymous** and **confidential**
- No personal identifiers (e.g. name, email, IP) will be collected
- Data will be stored securely and handled according to GDPR and university policies

Voluntary Participation:

- Participation is voluntary — you can withdraw at any time before submitting
- After submission, data is anonymized and cannot be withdrawn

Questions?

You can contact the researcher at yung-shiang.yang@ucdconnect.ie.

By clicking the button below, you confirm that:

- You are at least 18 years old
- You understand your participation is voluntary
- You may withdraw at any time before submission

I consent, begin the study

I do not consent, I do not wish to participate

Prolific ID

What is your Prolific ID? Please note that this response should auto-fill with the correct ID

${e://Field/PROLIFIC_PID}

Payment Information

💰 Payment Information

About ECU:
You'll earn and spend points called **ECU (Experimental Currency Units)**, **100 ECU = $1.0 USD**, this is used to calculate your bonus.

You will start with:
20 ECU + a chance to earn an additional **100 ECU** based on your decision and outcome in a randomly selected round in the experiment.

Make your decisions carefully.

Scenario Instructions - Control

📘 Scenario Overview

- You are applying for a job where a **hiring algorithm** decides who gets hired.

- There are two groups of applicants: **Green** and **Blue**. You've been assigned to the **${e://Field/GroupColor}**.

🎯 Goal of the Algorithm

- Each applicant will either become a **top** or an **average** performer in the job.
- The algorithm only wants to hire applicants who will become **top performers** in the job, but it can't tell who they are with certainty.
- Therefore, it asks that every applicant take a **pre-employment test** — a signal, though not a perfect one, of future performance.

📊 Algorithm's Belief

- The algorithm uses both your **test score** and your **group identity** to decide whether to hire you.
- It believes **Green** applicants are more likely to become top performers than **Blue** applicants — even though in reality, **everyone starts with a 50%** chance of becoming a top performer.

💡 Investment Option

- You can invest any amount from your base bonus of **20 ECU** to boost your chance of becoming **a top performer**.
- **1 ECU = +1.5%** boost. So, **10 ECU = +15%**; **20 ECU = +30%**.
- Your total chance of becoming a top performer = **50% + investment boost**.

🧪 The Pre-Employment Test

- If you're **a top performer**, you have 70% chance of scoring **A** (the highest score), 30% chance of scoring **B**, and 0% chance of scoring **C** (the lowest score).
- If you're **an average performer**, you have 0% chance of scoring **A**, 70% chance of scoring **B**, and 30% chance of scoring **C**.

🤖 The Algorithm's Hiring Rule

- **Green:** Hired if your score is **A or B**
- **Blue:** Hired only if your score is **A**

💰 Bonus

- **Base bonus:** 20 ECU
- **If hired:** +100 ECU
- **Important: Your investment cost is deducted from the base bonus whether you're hired or not.**

🔍 Examples

- Invest 10 ECU & hired → Bonus = 20 - 10 + 100 = **110 ECU**
- Invest 10 ECU & not hired → Bonus = 20 - 10 + 0 = **10 ECU**
- Invest 0 ECU & hired → Bonus = 20 - 0 + 100 = **120 ECU**

🧠 You'll complete 1 practice and 2 formal rounds. Your bonus depends on 1 randomly chosen formal round — so decide carefully!

Scenario Instructions - No Risk

📘 Scenario Overview

- You are applying for a job where a **hiring algorithm** decides who gets hired.
- There are two groups of applicants: **Green** and **Blue**. You've been assigned to the **${e://Field/GroupColor}**.

🎯 Goal of the Algorithm

- Each applicant will either become a **top** or an **average performer** in the job.
- The algorithm only wants to hire applicants who will become **top performers** in the job, but it can't tell who they are with certainty.
- Therefore, it asks that every applicant take a **pre-employment test** — a signal, though not a perfect one, of future performance.

📊 Algorithm's Belief

- The algorithm uses both your **test score** and your **group identity** to decide whether to hire you.
- It believes **Green** applicants are more likely to become top performers than **Blue** applicants — even though in reality, **everyone starts with a 50%** chance of becoming a top performer.

💡 Investment Option

- You can invest any amount from your base bonus of **20 ECU** to improve your chance of becoming a top performer.
- **1 ECU = +1.5%** chance. So, **10 ECU = +15%**; **20 ECU = +30%**.

- Your total chance of becoming a top performer = **50% + investment boost**.

🧪 The Pre-Employment Test

- If you're a top performer, you have 70% chance of scoring **A** (the highest score), 30% chance of scoring **B**, and 0% chance of scoring **C** (the lowest score).
- If you're an average performer, you have 0% chance of scoring **A**, 70% chance of scoring **B**, and 30% chance of scoring **C**.

🤖 Hiring Rule

- **Green:** Hired if your score is **A or B**
- **Blue:** Hired only if your score is **A**

💰 Bonus

- **Base bonus:** 20 ECU
- **If hired:** +100 ECU
- **Important: Your investment cost is only deducted from the base bonus if you're hired.**

🔍 Examples

- Invest 10 ECU & hired → Bonus = 20 - 10 + 100 = **110 ECU**
- Invest 10 ECU & not hired → Bonus = 20 - 0 + 0 = **10 ECU**
- Invest 0 ECU & hired → Bonus = 20 - 0 + 100 = **120 ECU**

🧠 **You'll complete 1 practice and 2 formal rounds. Your bonus depends on 1 randomly chosen formal round — so decide carefully!**

Attention Check

🧠 **Attention Check**
To ensure you've understood the instructions, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

Please drag and drop the following events into the correct order based on how the study works:

Test score is simulated

You are randomly assigned to a group

You decide how much to invest

The hiring algorithm makes a hiring decision

What is your chance of becoming a top performer in the job if you invest the maximum amount of 20 ECU?

(Hint: Everyone starts with a 50% chance of becoming a top performer. Each 1 ECU you invest increases your chance by 1.5 percentage points.)

65%

70%

75%

80%

Suppose, following the investment, you end up becoming a top performer. What are your chances of scoring A, B, C, respectively?

70%, 30%, 0%

0%, 70%, 30%

100%, 0%, 0%

Suppose you're a Green applicant with a score of B. Will you be hired?
*(Hint: The algorithm hires a Green applicant with either an **A** or **B**, only hires a Blue applicant with an **A**.)*

Yes

No

Suppose you're hired following your investment of 20 ECU. What will be the total amount of your bonus?

(Hint: You start off with a base bonus of 20 ECU, which you can invest at your discretion. You also get an additional bonus of 100 ECU if you're hired.)

120 ECU

100 ECU

0 ECU

20 ECU

Blue SRM

For your reference, before you make your investment decision, here is some information about previous Blue participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Blue	20	A	Yes
Participant B	Blue	20	A	Yes
Participant C	Blue	15	A	Yes
Participant D	Blue	10	A	Yes
Participant E	Blue	5	A	Yes

🧠 Attention Check

To ensure you’ve read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some

zero

The previous participants shown above were _____ by the algorithm.

hired

not hired

Green SRM

For your reference, before you make your investment decision, here is some information about previous Green participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Green	20	A	Yes
Participant B	Green	20	B	Yes
Participant C	Green	15	A	Yes

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant D	Green	12	A	Yes
Participant E	Green	11	A	Yes

🧠 Attention Check

To ensure you've read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some

zero

The previous participants shown above were _____ by the algorithm.

hired

not hired

Blue URM

For your reference, before you make your investment decision, here is some information about previous Blue participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Blue	20	B	No
Participant B	Blue	20	C	No
Participant C	Blue	15	B	No
Participant D	Blue	13	C	No
Participant E	Blue	10	C	No

🧠 Attention Check

To ensure you've read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some

zero

The previous participants shown above were _____ by the

algorithm.

hired

not hired

Green URM

For your reference, before you make your investment decision, here is some information about previous Green participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Green	20	C	No
Participant B	Green	11	C	No
Participant C	Green	10	C	No
Participant D	Green	10	C	No
Participant E	Green	10	C	No

🧠 Attention Check

To ensure you've read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some

zero

The previous participants shown above were _____ by the algorithm.

hired

not hired

Practice Round

🧪 **Practice Round**

This round is just for practice and will not affect your final bonus. It is designed to help you understand how the investment and test process work.

How much ECU would you like to invest to improve your probability of becoming a top performer? Please input below any whole number from 0 to 20.

Reminder: You are in the ${e://Field/GroupColor}. ${e://Field/GroupReminder}

> Your investment cost is deducted from your earnings whether you're hired or not.

Practice Simulation

Thanks for making your decision!
Your test result is now simulating...
⏳
Click ➡️ to reveal your results.

Practice Results

🧪 **Practice Round Result**

Here's how your decision played out in the practice round (this will not affect your final bonus):

💡 **You invested:** ${e://Field/PracticeInvestment}
🧠 **Your probability of becoming a top performer after investment is:** ${e://Field/PracticeProb}**%**
📊 **Your test score:** ${e://Field/PracticeTestScoreLabel}
🤖 **Hiring decision:** You are **${e://Field/PracticeHired}** by the hiring algorithm.
💰 **Your (hypothetical) bonus in the practice round:** 20 - ${e://Field/PracticeInvestmentDisplay} + ${e://Field/PracticeWage} = ${e://Field/PracticeEarnings} ECU

Note: This was just a practice round. You'll now complete **2 formal rounds**, and your bonus will be based on *one randomly selected formal round*.

Formal Round

✅ **Formal Round 1 of 2**

This is the first of two formal rounds. If this is the randomly chosen round, your decision here WILL affect your final bonus.

How much ECU would you like to invest to improve your probability of becoming a top performer? Please input below any whole number from 0 to 20.
Reminder: You are in the ${e://Field/GroupColor}.
${e://Field/GroupReminder}

Simulation

Thanks for making your decision!
Your test result is now simulating...
⏳
Click ➡️ to reveal your results.

Results

🎯 **Result of Round 1 of 2**

This was the first of two formal rounds. If this is the randomly chosen round, the outcome below WILL determine your final bonus.

💡 **You invested:** ${e://Field/Investment}
🧠 **Your probability of becoming a top performer after investment is:** ${e://Field/Prob}**%**
📊 **Your test score:** ${e://Field/TestScoreLabel}
🤖 **Hiring decision:** You are **${e://Field/Hired}** by the hiring algorithm.
💰 **Your bonus:** 20 - ${e://Field/InvestmentDisplay} + ${e://Field/Wage} = ${e://Field/Earnings} ECU

You've completed the first formal round. You'll now proceed to the second and final formal round.

Green SRM 2

For your reference, before you make your investment decision in the final round, here is some information about previous Green participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Green	20	A	Yes
Participant B	Green	20	B	Yes
Participant C	Green	15	A	Yes
Participant D	Green	12	A	Yes
Participant E	Green	11	A	Yes

🧠 **Attention Check**
To ensure you've read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some
zero

The previous participants shown above were _____ by the algorithm.

hired
not hired

Blue SRM 2

For your reference, before you make your investment decision in the final round, here is some information about previous Blue participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Blue	20	A	Yes
Participant B	Blue	20	A	Yes
Participant C	Blue	15	A	Yes
Participant D	Blue	10	A	Yes
Participant E	Blue	5	A	Yes

🧠 Attention Check

To ensure you've read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some

zero

The previous participants shown above were _____ by the algorithm.

hired

not hired

Green URM 2

For your reference, before you make your investment decision in the final round, here is some information about previous Green participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Green	20	C	No
Participant B	Green	11	C	No
Participant C	Green	10	C	No
Participant D	Green	10	C	No
Participant E	Green	10	C	No

🧠 Attention Check

To ensure you've read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some

zero

The previous participants shown above were _____ by the algorithm.

hired

not hired

Blue URM 2

For your reference, before you make your investment decision in the final round, here is some information about previous Blue participants who have completed this task:

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant A	Blue	20	B	No
Participant B	Blue	20	C	No
Participant C	Blue	15	B	No

Participant	Group	Investment (ECU)	Test Score	Hired?
Participant D	Blue	13	C	No
Participant E	Blue	10	C	No

🧠 Attention Check

To ensure you've read the information, please answer the following questions. (You will not be able to proceed unless you answer correctly.)

The previous participants shown above invested _____ ECU to improve their chances of becoming top performers.

some

zero

The previous participants shown above were _____ by the algorithm.

hired

not hired

Second Round No Risk Instructions

⚠️ Attention: Change in Rules

In the following round, your investment cost is **only deducted if you're hired**. Your investment cost is **not deducted if you're not hired**. The hiring rule from the hiring algorithm remains the same.

Second Round Control Instructions

⚠️ Attention: Change in Rules

In the following round, your investment cost **WILL be deducted whether you're hired or not**. The hiring rule from the hiring algorithm remains the same.

Second Formal Round

✅ Formal Round 2 of 2

This is the second and final formal round. If this is the randomly chosen round, your decision here WILL affect your final bonus.

How much ECU would you like to invest to improve your probability of becoming a top performer? Please input below any whole number from 0 to 20.

Reminder: You are in the ${e://Field/GroupColor}.
${e://Field/GroupReminder}

Second Simulation

Thanks for making your decision!
Your test result is now simulating...
⏳
Click ➡️ to reveal your results.

Second Results

🎯 **Result of Round 2 of 2**

This was the second and final formal round. If this is the randomly chosen round, the outcome below WILL determine your final bonus.

💡 **You invested:** ${e://Field/Investment2}
🧠 **Your probability of becoming a top performer after investment is:** ${e://Field/Prob2}**%**
📊 **Your test score:** ${e://Field/TestScoreLabel2}
🤖 **Hiring decision:** You are **${e://Field/Hired2}** by the hiring algorithm.

💰 **Your bonus:** 20 - ${e://Field/InvestmentDisplay2} + ${e://Field/Wage2} = ${e://Field/Earnings2} ECU

This concludes the decision-making part of the study. You'll now answer a short demographic questionnaire.

Demographics

Please answer the following short questions.
Your responses are anonymous and will be used for research purposes only. You will not be personally identifiable from your answers.

What is your age?
(Please enter in years)

What is your gender?

Male
Female
Non-binary / third gender
Prefer not to say

What is the highest level of education you have completed?

Some high school or less

High school diploma or GED

Some college, but no degree

Associates or technical degree

Bachelor's degree

Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, etc.)

Prefer not to say

What best describes your employment status?

Working full-time

Working part-time

Unemployed and looking for work

A homemaker or stay-at-home parent

Student

Retired

Other (please specify in the textbox below)

Please estimate your gross annual salary.

Less than 25,000 USD

Greater than/equal to 25,000 USD and less than 50,000 USD

Greater than/equal to 50,000 USD and less than 75,000 USD

Greater than/equal to 75,000 USD and less than 100,000 USD

Greater than/equal to 100,000 USD and less than 125,000 USD

Greater than/equal to 125,000 USD and less than 150,000 USD

Greater than/equal to 150,000 USD and less than 175,000 USD

Greater than/equal to 175,000 USD and less than 200,000 USD

Greater than/equal to 200,000 USD and less than 225,000 USD

Greater than/equal to 225,000 USD and less than 250,000 USD

Greater than/equal to 250,000 USD and less than 275,000 USD

Greater than/equal to 275,000 USD and less than 300,000 USD

Greater than/equal to 300,000 USD

Prefer not to say

In general, how willing are you to take risks?
0 = Completely unwilling to take risks 10 = Very willing to take risks

0 ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 ◯

Have you ever applied for a job or other opportunity where you believe your application was screened or reviewed by an AI system or algorithm?
(For example: resume filtering by applicant tracking systems used by companies and recruitment agencies; or credit checks by systems like FICO Score.)

Yes, I have submitted an application where I believe my application was screened by an AI system or algorithm

No, I have submitted applications but do not believe an AI system or an algorithm was ever used to screen them

I have never applied for a job or an opportunity that would involve screening

Not sure

Prefer not to say

If so, in which contexts have you interacted with the AI system or algorithm?
Select all that apply:

Job applications or hiring platforms

Credit scoring or loan applications

Other (please specify in the textbox below)

In general, how much do you trust major technology companies to make decisions that affect you <u>in a fair and unbiased way</u>?
0 = No trust at all 10 = Complete trust

0 1 2 3 4 5 6 7 8 9 10

In the experiment, what factor(s) below motivated your investment choice? Select all that apply.

How much previous participants invested

The potential bonus I could get if hired (100 ECU)

Whether previous participants were hired

My group identity (whether I was assigned to the **Green** or **Blue** group)

The starting amount I had available to invest (20 ECU)

Whether I have to pay for the investment cost when I'm not hired

Other (please specify in the textbox below)

Please let us know if you experienced any issue in the study, for example if anything was unclear or didn't work (optional).

End of Survey Message

Thank You for Participating!

This study investigates how people make decisions about investing in their skills when evaluated by a hiring algorithm. The hiring process in this experiment was simulated using fictional groups (**Green** and **Blue**) to study how perceived bias might affect decision-making. These group assignments were entirely hypothetical and not tied to any real-world identities.

Your bonus payment was calculated using Experimental Currency Units (ECU). The conversion rate is: **100 ECU = $1 USD**. The final bonus amount (if any) will be transferred to your Prolific account in the next few weeks.

If you have any questions about the study or would like to learn more about the research, feel free to contact the researcher at: yung-shiang.yang@ucdconnect.ie

Please click the button below to be redirected back to Prolific and register your submission. Alternatively, you can enter the following completion code to register your submission: **CS8K8PPI**