Experimental evidence on how implicit racial bias affects risk preferences

Daniel Auer

daniel.auer@carloalberto.org

and

Didier Ruedin

POST-PRINT

This is the final draft *after* refereeing

Published as: Daniel Auer & Didier Ruedin (2023) Experimental evidence on how implicit racial bias affects risk preferences, *Journal of Ethnic and Migration Studies*

https://doi.org/10.1080/1369183X.2023.2278396

Abstract

We ask how human behavior changes when racial discrimination is costly and when choices are risky. By asking N = 4,944 participants in Germany to form a soccer team in a series of online experiments, we measure decision-making in an accessible way. Higher costs of discrimination can reduce disparities, but we show that these costs can also trigger implicit racial bias: participants who received an additional financial incentive to select more skilled soccer players outperformed nonincentivized participants and differentiated less based on skin color. However, when confronted with risky choices in a lottery, incentivized participants are more likely to gamble to avoid players with a darker skin color. That is, racial (minority) markers alter the risk preferences of people when their decisions carry costly consequences. This implicit racial bias may partly explain why members of visible minority groups are regularly discriminated against in real-world competitive markets.

Experimental evidence on how implicit racial bias affects risk preferences *

DANIEL AUER DIDIER RUEDIN

Abstract

We ask how human behavior changes when racial discrimination is costly and when choices are risky. By asking N=4,944 participants in Germany to form a soccer team in a series of online experiments, we measure decision-making in an accessible way. Higher costs of discrimination can reduce disparities, but we show that these costs can also trigger implicit racial bias: participants who received an additional financial incentive to select more skilled soccer players outperformed nonincentivized participants and differentiated less based on skin color. However, when confronted with risky choices in a lottery, incentivized participants are more likely to gamble to avoid players with a darker skin color. That is, racial (minority) markers alter the risk preferences of people when their decisions carry costly consequences. This implicit racial bias may partly explain why members of visible minority groups are regularly discriminated against in real-world competitive markets.

Keywords: experiment, incentives, risk preferences, racial bias, discrimination.
JEL codes: C83, C99, D81, D91, J15.
Word count: 8,293

---

*__Daniel Auer (corresponding author)__: Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy and University of Mannheim, Germany; ORCID: 0000-0003-4454- 2365; daniel.auer@carloalberto.org. **Didier Ruedin**: University of Neuchâtel, Switzerland; ORCID: 0000-0001-5213-4316; didier.ruedin@unine.ch. **Acknowledgments**: We thank Thomas Tichelbaecker for outstanding research assistance. We are grateful to Rahsaan Maxwell and two anonymous reviewers for their guidance and excellent comments. For valuable input, we are grateful to Marc Helbling, Lena Hipp, Felix Jaeger, Kristin Kelley, Judit Kende, Sandra Morgenstern, Paula Protsch, Jonas Radl, Alyssa Taylor, Miri Song, Lieselotte Blommaert, Patrick Simon, the participants of the 2021 conference on Survey Experiments and Migration in Lausanne, the 2021 annual IMISCOE conference, and the 2021 WZB Experimentalists seminar. **Funding:** This research was supported by the WZB — Berlin Social Science Center, and by the National Center of Competence in Research *nccr — on the move* funded by the Swiss National Science Foundation under grant number 51NF40-182897. **Ethics:** The ethics assessment was approved by the Swiss Forum for Migration and Population Studies (SFM-20180313). **Preregistration:** This research was preregistered at https://aspredicted.org/fe48r.pdf.

1 Introduction

Making decisions is part of our daily routine. From trivial to the most complex choices, all decisions are influenced by risk preferences. The underlying risk assessments when making a decision typically occur subconsciously, for instance, when we choose a product in the supermarket. In general, dual-process theory regards decision-making as an interplay between deliberation and intuition (Epstein 1994). We argue that this interplay might also invoke racial discrimination when individuals are confronted with a decision-making process in which intuitive thinking overrides rationality (Butler et al. 2014). Therefore, we assess whether the skin color of candidates can alter individual risk preferences and subsequent behavior when monetary incentives trigger otherwise hidden implicit racial bias.

Our inference is based on a series of linked online experiments among a representative sample of 4,994 residents of Germany, in which participants had to choose soccer players to form a team and maximize its overall skill level. We used soccer as a frame to make the tasks more relatable to participants who may not necessarily be interested in politics.[1] Choosing among soccer players with clear guidelines on their skills entails fewer possibilities for consideration by the participant that remain unobserved to the researchers. The setting is also familiar to respondents, as soccer is extremely popular in Germany and beyond. Moreover, soccer players of different ethnicities and skin color enjoy reputations arguably more equal than those in most other domains of society. This ensures that we do not alienate participants and that — at most — our estimates of the prevalence of racial bias are conservative.

Our analyses demonstrate that participants who are randomly given a monetary incentive outperform nonincentivized participants in making informed decisions: monetary incentives encourage deliberation and a focus on objective skills, thus reducing discrimination when participants select among players with known skills (compare Levine et al. 2008; Kurzban et al. 2001). This effect is particularly notable for participants who ex post stated anti-immigrant attitudes. They do not discriminate against Black players when incentivized but do so in the absence of a monetary incentive. Importantly, this "competing away" of discrimination only occurs in settings with certainty over outcomes. By adding a risk component to the decision-making process, we remove the ability of participants to readily compare players. In a lottery

[1] Our approach is similar to studies such as Jepsen and Jepsen (2020), who examined behavior in a 'Fantasy Football League' among individuals with an interest in the sport and gaming.

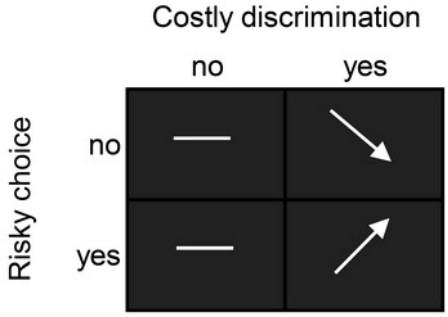
setting, we offer the choice between a player with known skills (the safe option) and a risky-option player who is either better or worse in terms of skills with $Pr = 0.5$ but has the same expected skill. The added risk encourages participants to rely on heuristics and gut feelings. This has the perverse consequence that monetary incentives encourage participants to opt for the risky option when the safe-option player has darker skin. In other words, the racial marker makes a certain alternative less attractive by adding a (subconscious) penalty, such that participants gamble more frequently. We argue that these findings point to deep-seated racial biases hidden within large parts of society. When decision-making is both consequential and risky, deliberation can succumb to intuition, and implicit racial biases influence choices.

In the context of diverse societies, finding behavioral differences between certain and risky choices opens a new perspective on how we understand racial discrimination: if risk preferences and subsequent behavior change with the racial, ethnic, or any other group marker of business partners, customers, vendors, or simply strangers on a bus, the combined role of stereotypes and risk preferences as an influence on individual behavior is dynamic, and — depending on the circumstances — incentives that would otherwise reduce or remove discrimination can actually trigger deeply ingrained drivers of it. In doing so, our study contributes to a body of literature that complements research on racial discrimination with insights from behavioral economics, political science, and social psychology (e.g., Bartoš et al. 2016).

2 Costly racial bias and risky choices

It is well established that some members of society have anti-immigrant attitudes and exhibit racial bias in all kinds of situations (Hainmueller and Hopkins 2014; Maxwell 2019; Pardos-Prado and Xena 2019). In market situations, anti-immigrant or more generally anti-minority attitudes often translate into discriminatory behavior. For example, members of minority groups receive fewer job offers and are paid less because of their group membership and not because they were individually different on relevant criteria such as human capital (Neumark 2018; Flage 2018). Here, we focus on two main channels through which the transition from negative stereotypes and anti-immigrant attitudes to discriminatory behavior can occur: the cost of discrimination and risky outcomes, summarized in Figure 1.

Figure 1: Changes in discriminatory behavior depending on risk and cost



Notes: The figure summarizes the hypothesized changes in discriminatory behavior depending on two contextual dimensions: risk and cost of discrimination. As indicated by the flat lines on the left, we assume no change in discrimination against racial minorities when the cost of discrimination is low (e.g., when behavior has no financial consequences). When the cost of discrimination is high and actors have certainty over competing outcomes, discrimination should decrease (top right). Conversely, discrimination should increase when choices are both costly and risky (bottom right).

2.1. Costly discrimination

Pager (2016) found that companies that discriminate against racial minorities in employment (i.e., hiring based on non-productivity-related racial preferences) are more likely to go out of business. Similarly, Levine et al. (2008) show that bank deregulation in the United States increased competition in the nonfinancial sector, which reduced the racial wage gap, especially in regions with high racial prejudice. Doleac and Stein (2013), in turn, found less price discrimination against Black vendors when demand-side competition was high. This implies that it should in principle be possible to "incentivize away" discrimination: if the cost of discrimination—i.e., the cost of foregoing the best of a set of N≥2 alternatives—increases sufficiently, discriminatory behavior should decrease. Translated to our present experimental setting, this means that if participants receive a monetary incentive to make the correct choice, we should see more effort to select the best alternative (i.e., to pick the soccer player with the highest skill level) and as a result observe lower levels of discrimination against Black players (Kurzban et al. 2001).

Hypothesis 1: Rewarding choices with monetary incentives reduces discriminatory behavior when decisions are made under certainty over outcomes.

2.2. Risky choices

The second main channel regards discrimination due to actors making risky choices. Taking a hiring situation as an example, employers do not know the exact skills of applicants, but they may derive them from signals such as diplomas or previous work experience. These signals are cheap proxies for costly efforts to assess the real skills of applicants (e.g., using a trial period in the company). If a rational employer knows the average productivity of a certain ethnic, racial, or any other group, membership status becomes a heuristic to quantify the risk of a job candidate having insufficient skills (Neumark 2018; Rivera 2020; Norris and Moss-Pech 2021). However, such signals also invite reliance on stereotypes and gut feelings, which can perpetuate ethnic and racial disparities, for instance, if an employer merely perceives members of a minority group to be less productive.

Using observational performance data in a sports setting, Norris and Moss-Pech (2021) show that Black basketball players in the United States are 30 percent more likely to drop out of the league than equally performing White players, indicating that racial disparities persist even when past performance indicators are readily available and when discrimination is costly (contract terminations on non-productivity-related grounds). One explanation for such findings is that risk (about future performance) switches the effect of the costly discrimination channel. Although we did not preregister this part of the analysis, we can make use of our experimental design, which simultaneously manipulates the cost of discrimination and the level of risk attached to different choices. We hypothesize that increasing the cost of discrimination through monetary incentives can, contrary to common expectations, *increase* discriminatory behavior when a choice is risky and when decision-makers hold deep-seated stereotypes against racial minorities. This *implicit racial bias* would mean that otherwise fruitful efforts to "incentivize discrimination away" outside the laboratory can result in more inequality when paired with risk over outcomes.

> Hypothesis 2: Implicit racial bias over the minority group's skills persists when choices are risky, such that monetary incentives increase discriminatory behavior.

3 Experimental setup

3.1 The task

We invited participants to complete an online survey of approximately 7 minutes with a fixed incentive of 0.5 euros. In total, 4,994 participants[2] were tasked to form a well-performing soccer team by adding a new player one-by-one in different experimental rounds.[3] The goal was to maximize the overall level of skill of the team to increase the chances of winning in a final game at the end of the survey. For this, two-thirds of the participants were offered an additional incentive of up to 1 euro.[4] Although 1 additional euro may sound small, it is twice the fixed incentive participants received for taking part.[5]

Theoretically, the skills could range from 0 to 99 (following common representations in video games), which we restricted to the average skills of the randomly drawn players s_i=[37; 98]. Figure 2 shows an example player, using a randomly drawn picture of a real soccer player and 4 randomly drawn skills. We chose real players because their portraits are standardized and all are dressed similarly (soccer jersey, athletic stature).

By having only male players, we provide homogeneity that allows us to better focus on the treatment. Note that we intentionally showed four dimensions and did not provide a player's overall skill level to render the correct choice (under skill maximization) less

[2] We drew the sample from 3.6 million users registered on the German online platform clickworker.com. Registered users are free to select tasks on the online platform conditional on being eligible (in our case, 18 years or older). Task descriptions typically contain a title (in our case "football academy"), the approximate duration, and the compensation amount. Our default incentive of 0.5 euros was above average compared with other tasks at the time. We do not know the total number of users who saw the survey invitation on the platform. The characteristics of the participants are summarized in Table A.1 in the Appendix. To minimize sampling bias, we invited participants in 11 waves with a target of 500 responses each starting in December 2019 (the beginning of soccer league break) and ending in March 2020 (the busy period in German soccer leagues). Based on unique IDs, we ensured that users could only participate once. While our sample represents the general population in Germany well, we acknowledge that the representativity of online surveys is generally limited.

[3] Participants were given 3 players at the start of the game for which we randomized two different average skill levels b = [73; 76]. These endowment players did not affect our subsequent analysis.

[4] One-third received a dynamic additional amount that was equivalent to the final average team strength *S* in euro cents $i_d = \overline{S}$. Given the average player skill level of approximately 71, randomly selecting a player every round resulted, on average, in an additional payoff of 0.71 euros for this treatment group. Another third received a fixed additional amount of 1 euro with probability equivalent to the final average team strength $i_f = 1 * \overline{S}$. Again, a random choice of eight players resulted in an additional payoff of $1 * 0.71 = 0.71$ euros on average. With this diversification, we sought to minimize the risk that participants misinterpreted the incentive payoff. However, both treatment groups showed identical behavior, such that we group them as 'incentivized participants'.

[5] We followed the guidelines of the platform in setting the average payoff, because paying substantially more (or less) can bias the sample.

obvious. However, we emphasized that all skill dimensions were equally important to calculate the average skill level.

Figure 2: Example of a player as presented to participants

Note: The figure shows an example of a player card presented to the participants. The picture was randomly scraped from a pool of real soccer players in the German third-tier league. Skills (physical STRength, MENtal strength, SHOoting, and PASsing) were randomly drawn from a distribution s_i = [37; 98] such that the average skill level of the players was $\overline{S} \approx 71$. See Figure A.1 for the overall skill distribution.

In total, participants added 8 players in 3 different types of games: 1 rating stage, 5 conjoint, and 2 lottery rounds. The experimental design is shown in Figure 3. Because the characteristics of the 3 endowment players were also randomized, their demographic makeup did not affect subsequent treatment effects.

Rating: In the first round, participants were asked to rate the skill of three players as a "warm-up". Figure A.2 in the Appendix shows that participants assess the true skill level (equally weighted average of all 4 dimensions) very accurately and without differentiating with respect to the players' skin color (which we explain in greater detail in Section 3.2). The average deviation for both groups was approximately −5.1 points, or 7 percent from the true average skill level. This is not very surprising, since the true

average skill level could have been calculated.[6] Participants were not told that the player they rated the most accurately was eventually added to their roster. Therefore, this task allows us to assess the general rating performance of participants without bias, so it can serve as a reference for the effect of monetary incentives on performance, as shown in Section 4.1.

Conjoint: In all subsequent rounds, participants were explicitly asked to select the best player, and at the beginning of every round, they were informed about the average strength of their current team and the goal to maximize it. Each performance round consisted of two one-in-five conjoint experiments (where participants were asked to select one out of five presented players) and two paired conjoint experiments each (where participants were asked to decide between two players). These tasks took place with certainty over outcomes. In other words, participants could, in principle, calculate the average skill level of each player and subsequently select the best player. On the basis of Hypothesis 1, we expect that discriminatory behavior decreases in the conjoint stages when participants are offered an additional monetary incentive to perform well.

Best choice task: After the conjoint rounds, participants were shown up to five players one by one, and for each player, participants had to decide whether to select the player or to be offered another player. This means participants had to decide whether a given player was `good enough' to be selected without knowing the skills of the subsequent players. Participants who forwent all four offered players were automatically assigned the fifth player to their roster.[7] We follow, for instance, Seale and Rapoport (1997) and use the best choice task as a proxy for risk preferences, assuming that risk-seeking participants are willing to forgo more (potentially good) players and gamble on (even) better players being revealed.

Lottery: Eventually, we asked participants in two lottery-style rounds to select either a safe-option player with a known skill level or a risky-option player who has a better or worse skill level with equal probability $Pr = 0.5$. This setting was explained transparently to the participants. We also randomly varied the possible gain/loss from 2

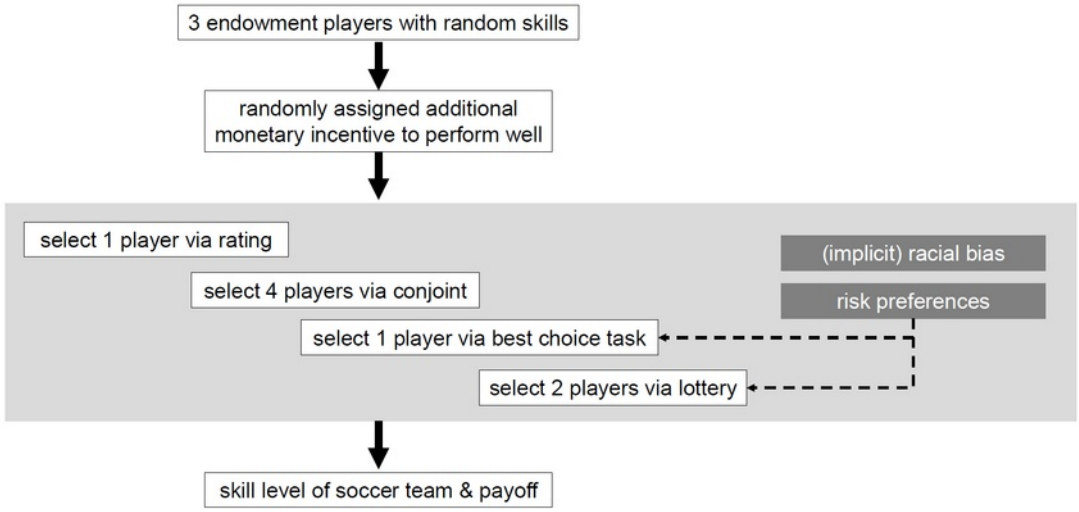
[6] To further test performance, we randomly left one of the four skill dimensions empty for 50 percent of the participants. As expected, this condition resulted in slightly weaker performance (average deviation with complete information = −3.9 vs. −6.3 with incomplete information). For the remainder, we ignore this informational treatment, as it had no effect on racial bias throughout the experiment.

[7] The median participant selected a player in the second round, that is, she or he had forgone 1 previous player (mean round of choice = 2.4). Studies frequently refer to this type of choice task as the "secretary problem" or "best choice problem" (e.g., Ferguson 1989), which belongs to the larger economic literature on optimal stopping (e.g., Kruse and Strack 2015).

to 12 skill points. For example, compared to a safe-option player with skills $\overline{S}=70$, participants could select a risky-player option with $\overline{S}=(68*0.5)+(72*0.5)$, or in an alternative scenario $\overline{S}=(58*0.5)+(82*0.5)$, and any possible combination in between. This lottery round reflects a common approach to elicit risk preferences (e.g., Dohmen et al. 2011; Charness et al. 2013).[8]

Final roster and payoffs: Once participants had formed their roster, they were informed about their potential additional payoff or—under the nonincentivized control—immediately transitioned to a general survey section with further sociodemographic questions. In total, we recorded N = 4, 994 unique participants who completed the survey and passed click time and plausibility checks to retain only valid answers.[9]

Figure 3: Experimental design



Note: The figure shows the experimental design and the two central causal mechanisms (racial bias and risk preferences; dark background on the right of the figure). All players presented to participants at all stages were fully randomized in terms of their skills and skin color. Participants were endowed with 3 players with given average skills (random variance) and random gray scores and were tasked with selecting 8 players in separate rounds of different games. Each round served a different purpose. First, participants rated players and were not aware that they would be assigned the player they rated most accurately (measure racial bias in a non-competitive setting). Second, participants picked 4 players in separate conjoint tasks (1 in 5 and 1 in 2) under certainty about skills and alternative players (measure racial bias and whether it can be "competed away"). Third, participants picked 1 player without knowing the skills of alternatives that would be offered in case the participant forwent the presented player (measure risk preferences). Fourth, participants picked 2 players by deciding between a safe and a risky option with equal expected quality (measure risk preferences and implicit racial bias).

[8] We henceforth only use the first lottery round for analyses because we observe that the probability of picking the safe player option in round 2 systematically depends on the mere choice in round 1 but not on the players' characteristics (choosing the safe option in round 1 reduces the probability of choosing the safe option in round 2 by 10 percentage points, independent of possible gain/loss and player traits).

[9] This number is smaller than what we preregistered, which decreases the likelihood of finding statistically significant effects. However, in consultation with the provider, we stopped data collection when the online panel was effectively exhausted.

3.2 Measuring skin color

The main goal of this experiment was to assess the role of skin color in participant behavior. A substantial literature reports systematic racial biases, resulting in discrimination against ethnic and racial minorities in competitive markets, such as the job, housing or rental market (e.g., Auer and Ruedin 2023; Doleac and Stein 2013; Quillian et al. 2020).[10]

Most studies that assess the role of racial markers in experiments present a limited set of profiles to their participants, such as Doleac and Stein (2013) who showed one Black hand and one White hand holding a product, or Gaddis (2014) who sent fake job applications in the United States using 3x2 names carrying a racial signal. Here, we take a more general approach that is fully transparent and cannot be biased by subjectivity: we calculated a *gray score* of each player picture we presented to the participants in the experiment. This approach also allowed us to construct a metric measure of skin color, similar to but more detailed and less subjective than, for example, the NIS skin color scale as proposed by the New Immigration Survey in the United States (c.f. Massey 2011).[11] Another advantage is that we were able to field many photographs (N = 575), meaning that the results could not have been biased by small-sample issues.

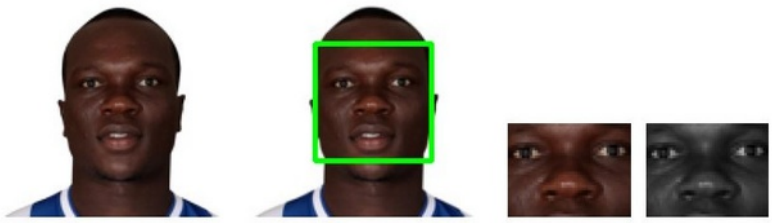
We scraped and randomly selected standardized portraits of 575 real soccer players in the German third-tier league. This sample ensures realism (male athletes in typical pose) and facial homogeneity among portraits of different skin colors (expressions such as smiling, etc.). At the same time, using third-tier league players minimizes the likelihood of players being recognized by their mere portrait (which would be an issue when using pictures of renowned first-league players). Because we do not provide names, recognition of a randomly sampled portrait is extremely unlikely: for instance, we can approximate the probability that a respondent recognizes a player of their favorite team. Therefore, we take the total number of social-media followers from different platforms of the team with the most followers (534,000 followers of 1.FC Kaiserslautern, c.f. liga3-online 2015). Assuming these were unique supporters (which is unlikely), the

[10] Such patterns of discrimination are not limited to skin color but are also prevalent for other markers, such as ethnicity (Quillian et al. 2017), wearing a headscarf (Weichselbaumer 2020), or nationality (Auer and Fossati 2019).

[11] We focus on skin color and neglect other physiological markers that we cannot measure systematically but may invoke racial bias

probability that we sampled a supporter of the most popular team from the German population (0.6%) and that this supporter was shown the portrait of a player of her favorite team—whom she presumably recognized—was 0.9% (28 of 575 players times a maximum of 29 players shown in the survey). Hence, it is safe to assume that respondents recognized the typical traits of a soccer player (athleticism, posture, etc.), but could not identify team affiliation, nationality, or name of a single player.

Figure 4: Measuring skin color using a gray score



Note: The figure shows the creation of the players' *gray score* by calculating the average (mean) gray score (0 − 255) of the pixels covered by the green frame.

Figure 4 shows how we calculated the gray score of the portraits. The standardized size and aspect ratio of these portraits—similar to passport photographs—allowed us to define a frame that is covered by the player's face in every picture (the player's hair is not included to focus on skin color). Each frame was then transformed into a simple grayscale picture: A pixel can take 256 different intensities of gray color, with 0 being white and 255 being black. With this information, we calculated the average *gray score* of all pixels in the frame, which we eventually standardized such that $\overline{G} = [0...1]$.[12]

4 Results

4.1 Incentives and performance

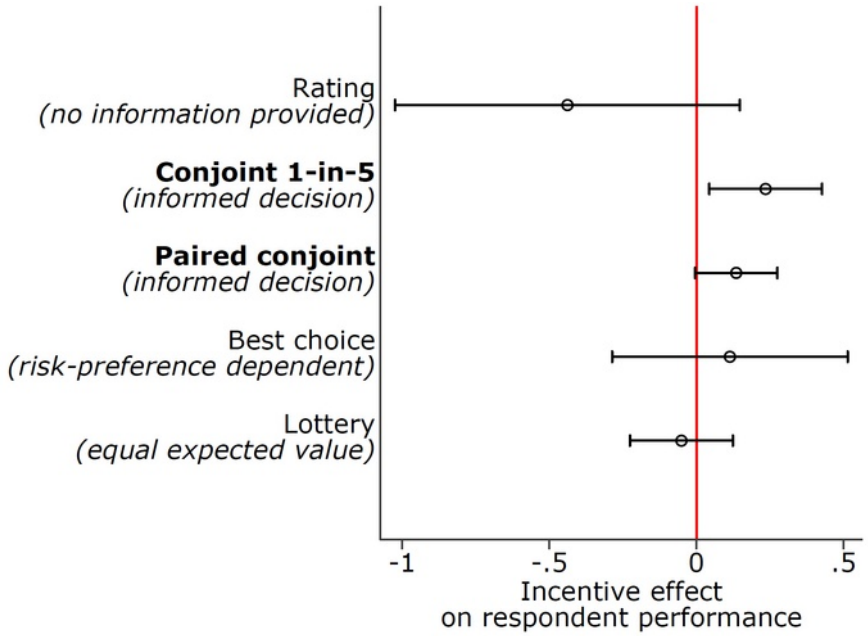
We begin by testing the effect of receiving a monetary incentive on performance during the experiment. To do so, we estimate:

[12] We qualitatively tested the validity of these scores by comparing them with the assessments of student assistants, yielding high correspondence. Moreover, our research assistants coded the facial expressions shown in the pictures. To avoid biased assessments due to skin color, we used grayed-out versions of the players' pictures and found no correlation between skin color and smile, friendliness, or subjective beauty.

$$S_{ik} = \lambda_k + \gamma\left(k * T_i\right) + X_i^{'}\delta + \varepsilon_{ik}, \quad (1)$$

where the skill level *S* of the player chosen by participant *i* in round *k* is a function of the baseline effect of game round *k*, interacted with a binary indicator for whether the participant was offered a performance-based additional incentive *T*. We also adjusted for the participant's age, gender, and education *X*. Figure 5 presents the coefficient *γ*, that is, the effect of the monetary incentive on the chosen players' skill level for each round. The top coefficient plots the *rating round* : participants did not know that a player would be added to their roster ("warm-up"), so we did not expect any effect on participant performance. In the *best choice task*, player skill levels were randomly distributed, such that incentivized participants could not outperform their nonincentivized counterparts. We use this task as a proxy for risk preferences. The same holds for the *lottery* round, in which the expected payoffs of the risky option and the safe option are identical. Offering a monetary incentive that depends on participant performance (maximize team strength) only results in a statistically significant selection of better players in the conjoint stages, the only rounds in which participants could effectively influence skill levels beyond chance. Hence, Figure 5 confirms that incentives boost performance. Figure A.3 in the Appendix further shows that it took incentivized participants significantly longer, on average, to complete the survey, indicating that they gave more thought to their decisions.

Figure 5: Incentivized participants outperform nonincentivized counterparts in tasks where they could do so



Note: The figure shows the effect of a monetary incentive on the skill level of the chosen player (DV) at each round of the experiment. Participants were not informed during the *rating* round and, by construction, had no possibility of outperforming nonincentivized participants during the *best choice* and *lottery* rounds. At the *conjoint* stages (bold labels), incentivized participants chose, on average, a player with a skill level 0.2 points better. OLS point estimates with 95% confidence intervals.

4.2 Incentives and discrimination under certainty

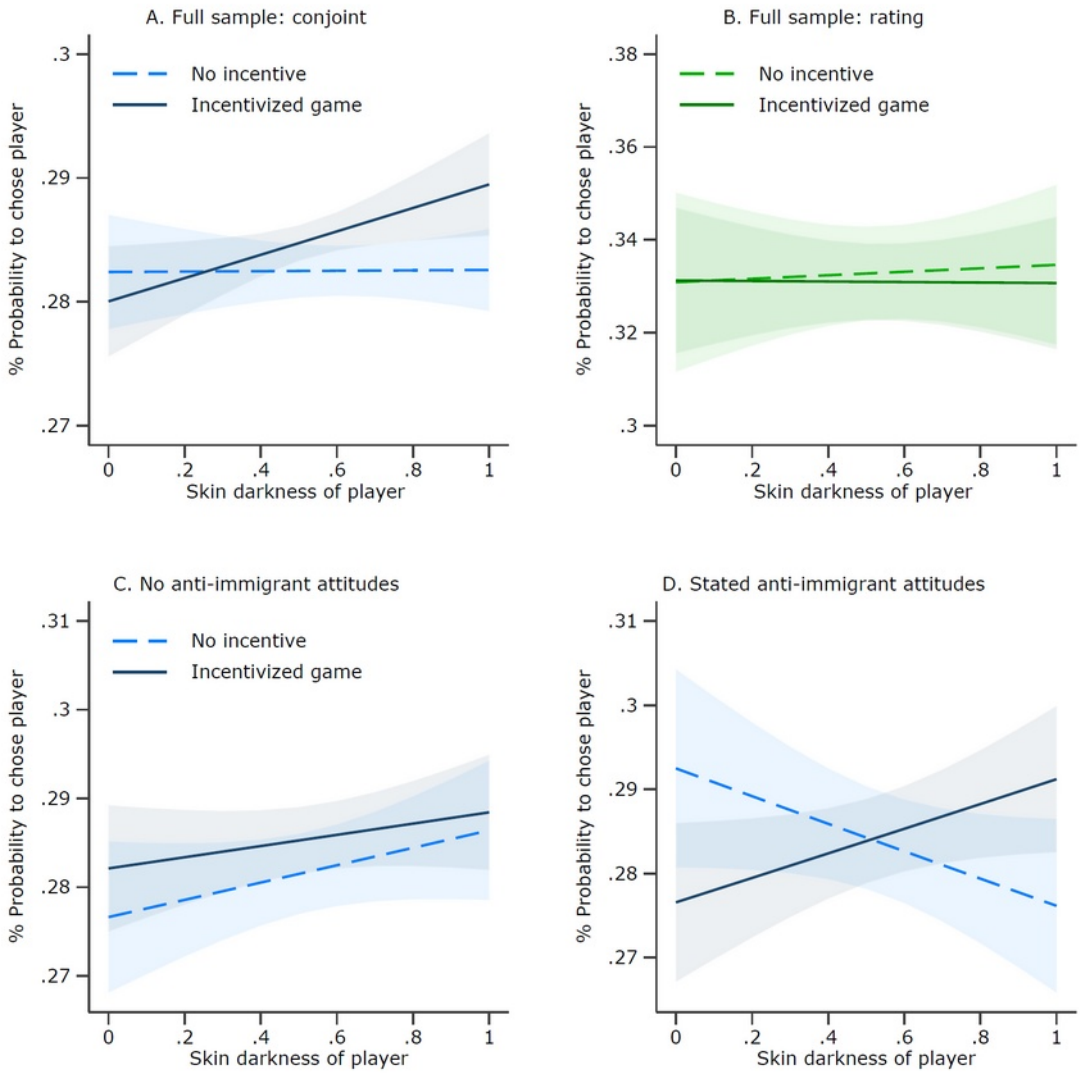
Next, we assess whether the performance increase shown in Section 4.1 is indeed related to discrimination being "competed away" (H1). Therefore, we focus on the two conjoint stages, which represent the experimental rounds in which participants have full control over the outcome (which player is chosen) and certainty over outcomes (the skill levels of the players up for selection). We estimate the following logit model:

$$Pr\left[Y_{pik}\right] = \tau\left(G_p * T_i\right) + X_i^{'}\delta + \lambda_k + \varepsilon_{pik}, \qquad (2)$$

where the probability of participant *i* selecting player *p* is a function of the player's gray score *G* (skin color) interacted with an indicator *T* for whether the participant received a monetary incentive. Subsequently, we adjusted for the player's skills (shown to the respondent) and participant characteristics (age, gender, and education). Due to the fully randomized nature of our experiment, if anything, adding controls should add precision to the point estimates. Furthermore, we adjust for round fixed effects *k* because the

baseline probabilities differ between the 1-in-2 and the 1-in-5 selection rounds and may also vary within each conjoint setup. Figure 6.A plots the predicted probabilities of picking a player. Conditional on the player's observed skills, participants who were incentivized to perform well had a 7 percent higher probability of picking a Black player, on average. This effect is statistically significantly different from zero. The corresponding regression Table A.2 in the Appendix also confirms that the results are robust to adding controls. As a robustness check, we show that, as expected, gray score has no effect on the rating in the previous round and that, consequently, there is no heterogeneity between incentivized and nonincentivized participants (Figure 6.B).

Figure 6: Incentives alter behavior



Note: Predicted marginal probability of choosing a player under certainty, conditional on the player's skin color (gray score). Margins are estimated at intervals of [0.1]. Logit point estimates and 90% confidence intervals.

Because we did not expect every participant to change behavior in response to skin color, we probed which part of the sample was driving the results of the conjoint rounds. Therefore, we split the sample along a question that was asked in the survey's post-treatment battery on attitudes. Participants had to state their (dis)agreement with the statement that natives should be favored over immigrants in the labor market, which is a standard question in established surveys such as the World Values Survey or the European Social Survey. We use native preference as a proxy measure for anti-immigrant and racist attitudes, as studies have frequently shown that racial stereotypes and anti-immigrant attitudes are closely linked (e.g., Schindler and Westcott 2021).[13] Figure 6.C shows that the probability of picking a player slightly increases with gray score for both incentivized and nonincentivized participants. In contrast, the heterogeneous effect of a monetary incentive is driven by participants who explicitly stated racist attitudes (Figure 6.D). While nonincentivized participants who agree with a racist statement have a substantially lower probability of picking a Black player, incentivized participants who agree with the statement behave similarly to their nonracist counterparts. This heterogeneity confirms the hypothesis that discriminatory behavior reacts to incentives, and, as such, can be competed away (Becker 1957; Kurzban et al. 2001).

As a sensitivity check, we can also restrict to players who were chosen by participants during the conjoint rounds and specify a simple OLS model, where the gray score of a player *i* chosen by participant *p* is a function of whether the participant was incentivized:

$$G_{pik} = T_i \psi + X_i^{'} \delta + \lambda_k + \varepsilon_{pik}, \tag{3}$$

Again, we leverage full randomization of the incentive treatment, the players' skin color, and the players' skills. Under a no-discrimination scenario the coefficient of the

[13] We do not have a question on skin color and rely on the empirical correlation between attitudes toward immigrants and racial minorities (Kalkan et al. 2009; Ruedin 2020) in a context where questions on skin color are heavily influenced by social-desirability bias (Creighton et al. 2015). We define the indicator variable for anti-immigrant attitudes as 1 if the participant (somewhat) agrees with the question (translated from German: "Employers should favor Germans over immigrants when jobs are scarce."), and 0 if the participant (somewhat) disagrees. Note that the results are robust to more restrictive definitions of this indicator variable. Figure A.4 in the Appendix plots the distribution of (dis)agreement with the statement.

incentive would be 0, meaning that participants are indifferent about the players' skin color and consequently do not alter their choice behavior when incentivized. Model 6 in Table A.2, however, shows that the average skin color of the chosen players is significantly darker when respondents have a monetary incentive to pick the best players.

4.3 Incentives and discrimination under risk

Thus far, participants could make an informed decision under certainty. Confirming Hypothesis 1, we find that adding a cost to discriminatory behavior—participants not picking the best player and thus risking not receiving the additional reward—reduces discrimination. Hypothesis 2 refers to the dynamics of discriminatory behavior when participants choose under risk. To test it, we restrict the sample to the lottery round with equal expected payoffs. A risk-neutral participant is indifferent between choosing the safe-option player (known skill level) or the risky-option player (equally lower or higher skill level with Pr = 0.5 each). Human biases may play against this, notably the risk-averse tendency in decision-making (i.e., most participants are expected to choose the safe option; e.g., Schildberg-Hörisch 2018). In any case, we should observe no effect of skin color. However, if racial bias is prevalent among the population (e.g., Norris and Moss-Pech 2021), skin color may alter the perceived relative payoffs of the two options, such that subjective beliefs about the expected skill level of a player with a lighter skin color are more optimistic than those about a player with a darker skin color. In other words, racial bias might add a skill penalty to Black players, making risky-option White players relatively more attractive. As a consequence of racial bias that—subconsciously—affects risk preferences, conscious effort to focus more on skills when incentivized to do so may be counteracted.

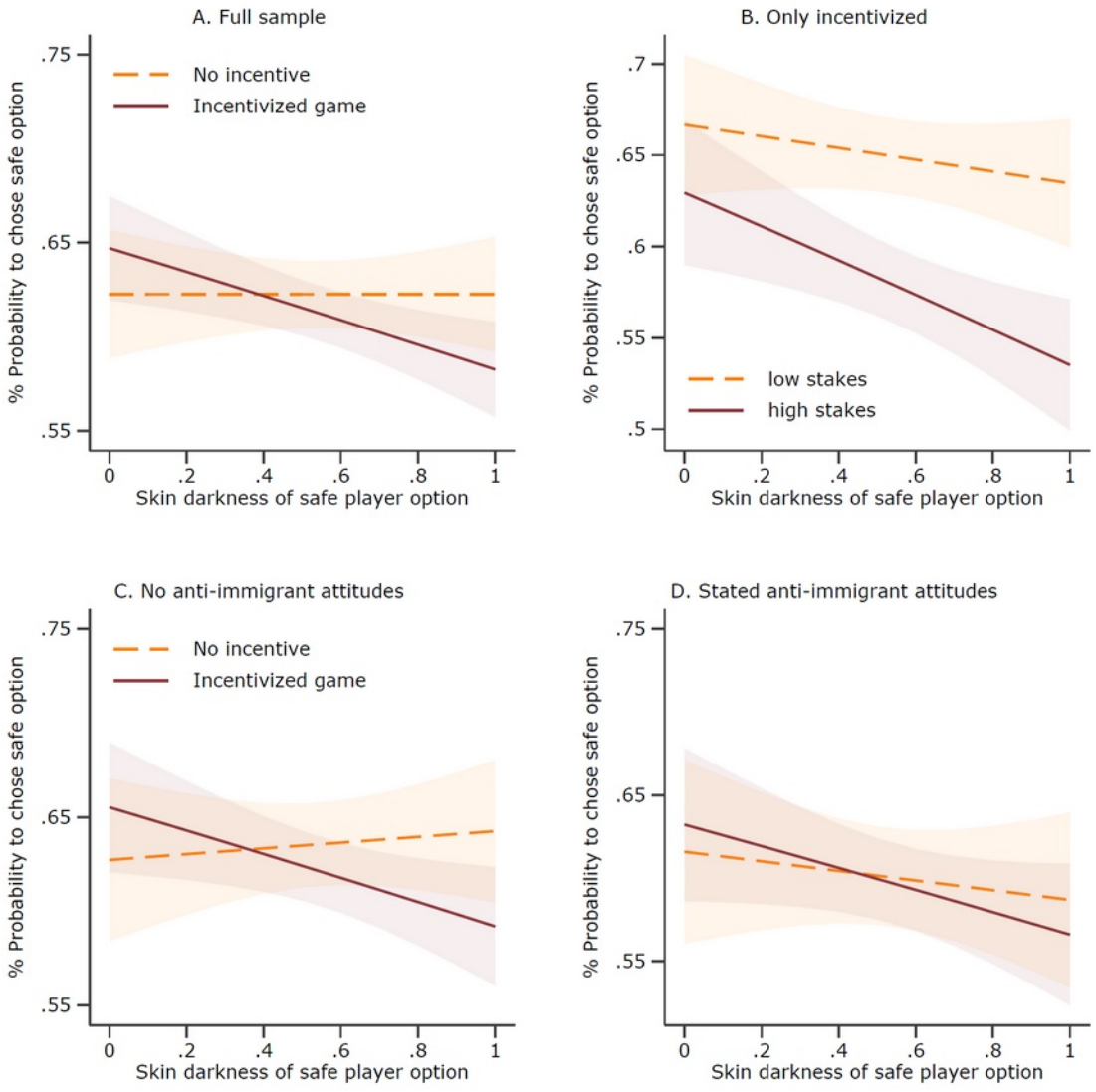
Figure 7 plots the predicted probabilities of picking the safe-option player by the participant's monetary incentive status and the gray score of the shown safe player. The predicted margins are estimated using the following logit model:

$$Pr\left[Y_{ipar}\right] = \tau\left(G_p * T_i\right) + X_i^{'}\delta + \psi_a + \phi_r + \varepsilon_{ipar}, \tag{4}$$

which essentially reflects Equation 2. Here, we need not add game round fixed effects, but instead adjust for the participant's decision in the preceding best choice task (ψ_a

ranging from 1 to 5, to condition on individual risk preferences)and the potential gains/losses of the risky-option player (φ_r , +/–2 to +/–20 skill points). The corresponding regression table is shown in Appendix Table A.3.

Figure 7: Skin color and risk preferences by anti-immigrant attitudes



Note: Predicted marginal probability of choosing the safe-option player in the lottery stage, conditional on the player's skin color (gray score). Margins are estimated at intervals of [0.1]. Logit point estimates and 90% confidence intervals. Figure A.6 in the Appendix shows additional heterogeneity tests for gender, education, and a pretreatment measured proxy for risk preferences. Both men and women become more risk seeking and converge to risk neutrality when incentivized and confronted with a Black player as the safe option. Furthermore, the gap between incentivized and nonincentivized behavior is stronger for low-educated and relatively younger (below 33 years) participants, whereas skin color does not make a substantial difference for highly educated or older participants, regardless of whether they are incentivized. Ultimately, incentives mainly change the preferences of risk-averse participants.

For the nonincentivized sample in Figure 7.A, the skin color of the safe-option player has no effect on the probability of choosing this risk-averse alternative, so the predicted probability remains constant at approximately 62 percent. This share closely reflects the risk preferences of the general population (Schildberg-Hörisch 2018). For participants who were offered the performance-based incentive, there is a clear downward slope, resulting in a predicted probability of choosing the safe option when the player has Black skin color of approximately 58 percent. The individual controls agree with general differences in risk evaluations (e.g., Dohmen et al. 2017; Schildberg-Hörisch 2018), which validates our approach: relatively older and female participants are more risk averse, whereas higher potential gains from selecting the risky option (risk level) and a higher stage in the previous best choice task (gambling for a better player to come) are associated with reduced risk aversion in the lottery round. We find a small positive effect of education on picking the safe option, while participants who hold anti-immigrant attitudes seem to be more risk-seeking. Importantly, this racial penalty cannot be rationalized by incomplete information.

Figure 7.B restricts to incentivized participants only but splits the sample along the randomly assigned potential gains/losses of the risky-option player. For readability, we created a binary variable indicating whether stakes were relatively low (50% chance that the risky player's skills are 2 to 10 points higher/lower) or relatively high (12 to 20 points higher/lower). In both scenarios, incentivized participants became more reluctant to choose the safe-option player if his skin color was darker. However, the change is significantly stronger when stakes are high, which further supports the hypothesis that implicit racial bias alters risk preferences with increasing economic consequences.

In Figure 6 we have shown that when deciding under certainty over outcomes, the behavioral change of a monetary incentive was stronger among participants who expressed anti-immigrant attitudes. We again test for this heterogeneity by plotting the group-specific probabilities of picking the safe-option player during the lottery round. Among the incentivized participants, those who did not articulate anti-immigrant sentiments (Figure 7.A) react very similarly to their counterparts with a more restrictive stance on immigration (Figure 7.B). In fact, the slope of the skin color effect is almost identical. However, for nonincentivized participants, the slopes point in opposite directions. This contrariwise heterogeneity offers two conclusions. First, without monetary incentives, people act upon their explicit preferences, meaning that those who stated anti-immigrant preferences are also less likely to stick with a racialized (Black)

player when the opportunity to swap for a White player occurs. Conversely, nonincentivized participants without stated preexisting anti-immigrant attitudes also did not discriminate based on skin color. Second, when performance is incentivized (i.e., failure to perform results in economic loss), previously neutral participants fall victim to implicit racial bias and converge to participants with explicit anti-immigrant stereotypes. That is, we observed no changes in risk preferences for participants with anti-immigrant stereotypes when a monetary incentive to perform the task well is present—they discriminate against Black players in both states, incentivized or nonincentivized. However, participants without explicit anti-immigrant stereotypes significantly changed their behavior, such that incentivized participants now take the chance to swap a safe Black player for a risky (White) player. Again, we do not observe this pattern when decisions are made under certainty. Together, this suggests that people implicitly ascribe an additional uncertainty or skill penalty to racial (or ethnic) minorities.

The random assignment of the gray score and skill level to the safe and the risky options means that we need not adjust for the characteristics of the risky player in Equation 4. We test this conjuncture by estimating the choice probability of all players in the lottery and adding an indicator variable to determine whether the player option is safe or risky. The coefficients in Model 5 of Table A.3 confirm our main results and add additional precision to the point estimate of the interaction term.

In sum, the results support Hypothesis 2 that discrimination increases and implicit racial bias penalizes Black players when decision-making is both risky and costly (Figure 7). In the Appendix, we provide a series of additional sensitivity checks, all of which confirm our findings. In Table A.4, we split the sample according to incentive treatment status and added further controls. The penalty ascribed to safe players with darker skin color is now more precisely estimated and becomes statistically different from zero. We also show that the results do not change when adjusting for (endogenous) performance of the participants prior to the lottery stage. The results are also robust to different sample restrictions (excluding speeders/dawdlers and non-German citizens, Figure A.5). Finally, we re-estimate Equation 2 and replace the gray score of the safe-option player with the gray-score distance between the safe-option player and the risky-option player (Table A.5), again confirming our hypothesis that racial markers carry a perceived performance penalty and that this implicit racial bias manifests itself in consequential market situations.

5 Discussion and Conclusion

We analyze discriminatory behavior when the cost of discrimination and the risk over outcomes vary simultaneously. By experimentally altering these two common drivers of discrimination, we show that the combined role of stereotypes and risk preferences is *dynamic*. To reach this conclusion, we first verify that discrimination can be "competed away" (e.g., Becker 1957; Levine et al. 2008). In other words, incentives to perform better in terms of skill maximization render discriminatory behavior a costly alternative, which leads to more careful attention to skills and less focus on skin color. This mechanism applies in situations featuring certainty over outcomes, where better players could readily be identified.

In the lottery stage of the experiment, we study how decision-making changes when simultaneously adding a risk component. Without a monetary incentive, participants do not deviate from expected patterns and predominantly behave in a risk-averse manner. Incentivized participants, however, tended to engage in more risk-*seeking* behavior when the safe option was a Black player. In other words, the presence of darker skin affects the relative risk evaluation between two alternatives. Racial biases that are incentivized away when participants have certainty over outcomes reemerge under risk. By demonstrating that participants are more willing to gamble when the (safe) alternative means being 'stuck' with a Black player, we highlight that risk preferences are probably more dynamic and amenable to context than generally assumed. We show that risk preferences not only vary with age and gender (e.g., Dohmen et al. 2017; Schildberg-Hörisch 2018) but are also influenced by the characteristics of available options that are unrelated to their potential payoffs (skin color, as opposed to human capital, for instance).

The heterogeneity analysis revealed that these effects were primarily driven by participants with low education, high risk aversion, and younger age. For others, such as highly educated participants, social norms and a focus on deliberation over intuition can overshadow the underlying racial prejudice (Kunstman et al. 2013), whereas those who openly expressed anti-immigrant attitudes always tended to discriminate against Black players. A possible implication of our result is that we should make diversity a goal in itself—alongside the goal of selecting the best candidates. In other words, incentives need to be designed to reinforce antiracist norms (e.g., Blinder et al. 2013; Kunstman et

al. 2013) and not to (implicitly) act against them. Future research should investigate in greater detail which individual characteristics drive our heterogeneous effects and in which groups of society social norms are most likely to succumb to implicit racial bias.

We acknowledge additional avenues for future research. Specifically, it is possible that variation in the non-economic context affects whether and to what extent risk preferences change. For instance, people might attempt to "homogenize" or "diversify" their team based on skin color. Here, we only have limited leeway to evaluate, given that we have to assume that the choices made during the experiment are endogenous. The additional sensitivity analyses show that controlling for the (endogenous) team composition in terms of average quality and gray score does not affect the probability of picking a safe-option player, on average. Moreover, we find some heterogeneity by economic pressure in the lottery round (i.e., high vs. low stakes). In general, making assumptions about the fundamental psychological functioning of decision-making is beyond the scope of this study. However, we find support for our argument in research at the intersection of economics and psychology. For instance, Butler et al. (2014) show that "intuitive thinkers tolerate more risk [...] than effortful reasoners". Future research should attempt to uncover the psychological mechanisms underlying the relationship between risk preferences and subconscious racial bias.

Methodologically, we demonstrate that we can also derive racial discrimination from more "approachable" settings than, for instance, confronting participants with highly abstract hypothetical hiring situations. McDonald (2019) explicitly warns against using abstract, hypothetical situations because they lack realism and are cognitively difficult to process for survey participants. We readily admit that most of our participants have probably never chosen a real soccer team. However, given that soccer is by far the most popular sport in Germany (87 percent in our sample are at least 'a little interested', which is comparable to results from market research, and 83 percent watch soccer games) and our participants play sports video games an average of 1.2 days per week, we argue that they are more likely to be familiar and comfortable with this setting. In this sense, participants can relate to the task, so we can better capture the underlying mechanisms of racial discrimination and decision-making under risk. With the focus on soccer, we also chose a case in which darker skin is not necessarily only subject to negative stereotypes, since there are narratives that equate Blackness with athleticism (Sailes 1993), and soccer arguably ranks among the most diverse parts of society. The implication is that our results are conservative (compare Jepsen and Jepsen 2020).

Research in different contexts is needed, but situations outside sports may struggle to provide a credible setting where the necessary skills can be measured in a comparably 'objective' way (Norris and Moss-Pech 2021).

We argue that implicit racial bias—as opposed to deliberate racism or xenophobia—is more widespread and affects behavior at the subconscious level. Our results imply that empirical research on racial and ethnic discrimination should focus more on risk preferences and how they influence individual behavior (see also Jamieson et al. 2013). Thus, we join recent calls for increased scrutiny of employers and decision-makers in the study of discrimination (Rivera 2020; Di Stasio and Lancee 2020). Specifically, our results suggest that racial biases can reemerge when assessing risky outcomes, implying that competitive markets and everyday situations trigger stereotypes and discriminatory behavior. These findings possibly suggest a novel explanation for the racial/ethnic inequalities that persist in most societies (e.g., Dovidio et al. 2010; Chetty et al. 2020). Future research should more carefully consider the dependency of risk preferences and racial bias in everyday interactions to better understand how they simultaneously shape discriminatory behavior.

References

Auer, Daniel, and Flavia Fossati. 2019. “The absent rewards of assimilation: how ethnic penalties persist in the Swiss labour market.” *Journal of Economic Inequality* 17:285–99. https://doi.org/10.1007/s10888-018-9401-2.

Auer, Daniel, and Didier Ruedin. 2023. “How one gesture curbed ethnic discrimination.” *European Journal of Political Research* 62 (3): 945–966. https://doi.org/10.1111/1475-6765.12547.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. “Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition.” *American Economic Review* 106 (6): 1437–75. https://doi.org/10.1257/aer.20140571.

Becker, Gary S. 1957. *The economics of discrimination.* Chicago: University of Chicago Press.

Blinder, Scott, Robert Ford, and Elisabeth Ivarsflaten. 2013. “The Better Angels of Our Nature: How the Antiprejudice Norm Affects Policy and Party Preferences in Great Britain and Germany.” *American Journal of Political Science* 57 (4): 841–57. https://doi.org/10.1111/ajps.12030.

Butler, Jeffrey, Luigi Guiso, and Tullio Jappelli. 2014. “The role of intuition and reasoning in driving aversion to risk and ambiguity.” *Theory and decision* 77:455–484.

Charness, Gary, Uri Gneezy, and Alex Imas. 2013. “Experimental methods: Eliciting risk preferences.” *Journal of Economic Behavior & Organization* 87:43–51. https://doi. org/10.1016/j.jebo.2012.12.023.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020. “Race and Economic Opportunity in the United States: an Intergenerational Perspective.” *The Quarterly Journal of Economics* 135 (2): 711–783. https://doi.org/10.1093/qje/qjz042.

Creighton, Mathew J., Amaney Jamal, and Natalia C. Malancu. 2015. “Has Opposition to Immigration Increased in the United States after the Economic Crisis? An Experimental Approach.” *International Migration Review* 49 (3): 727–756. https://doi.org/10.1111/imre.12091.

Di Stasio, Valentina, and Bram Lancee. 2020. “Understanding why employers discriminate, where and against whom: The potential of cross-national, factorial and multi-group field experiments.” *Research in Social Stratification and Mobility* 65:100463. https://doi.org/10.1016/j.rssm.2019.100463.

Dohmen, Thomas, Armin Falk, Bart H. H. Golsteyn, David Huffman, and Uwe Sunde. 2017. “Risk Attitudes Across The Life Course.” *The Economic Journal* 127 (605): F95–F116. https://doi.org/10.1111/ecoj.12322.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Ju¨ rgen Schupp, and Gert G. Wagner. 2011. “Individual risk attitudes: measurement, determinant, and behavioral consequences.” *Journal of the European Economic Association* 9 (3): 522–550. htt ps://doi.org/10.1111/j.1542-4774.2011.01015.x.

Doleac, Jennifer L., and Luke C.D. Stein. 2013. “The Visible Hand: Race and Online Market Outcomes.” *The Economic Journal* 123 (572): F469–F492. https://doi.org/10.1111/ecoj.12082.

Dovidio, J., M. Hewstone, P. Glick, and V. Esses, eds. 2010. *The SAGE Handbook of Prejudice, Stereotyping and Discrimination.* Thousand Oaks: Sage. ISBN: 1-4129-3453-2.

Epstein, Seymour. 1994. “Integration of the cognitive and the psychodynamic unconscious.” *American psychologist* 49 (8): 709.

Ferguson, Thomas S. 1989. “Who solved the secretary problem?” *Statistical science* 4 (3): 282–289.

Flage, Alexandre. 2018. “Ethnic and gender discrimination in the rental housing market: Evidence from a meta-analysis of correspondence tests, 2006–2017.” *Journal of Housing Economics* 41:251–273. https://doi.org/10.1016/j.jhe.2018.07.003.

Gaddis, S. Michael. 2014. “Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market.” *Social Forces* 93 (4): 1451–1479. https://doi.org/10.1093/sf/sou111.

Hainmueller, Jens, and Daniel J. Hopkins. 2014. “Public Attitudes Toward Immigration.” *Annual Review of Political Science* 17 (1): 225–249. https : / / doi . org / 10 . 1146 / annurev-polisci-102512-194818.

Jamieson, Jeremy P., Katrina Koslov, Matthew K. Nock, and Wendy Berry Mendes. 2013. “Experiencing Discrimination Increases Risk Taking.” *Psychological Science* 24, no. 2 (February): 131–139. https://doi.org/10.1177/0956797612448194.

Jepsen, Christopher, and Lisa Jepsen. 2020. “Race and National Football League Player Salaries After Controlling for Fantasy Statistics and Arrests.” *Journal of Sports Economics.*

Kalkan, Kerem Ozan, Geoffrey C. Layman, and Eric M. Uslaner. 2009. ““Bands of Others”? Attitudes toward Muslims in Contemporary American Society.” *The Journal of Politics* 71 (03): 847–862.

Kruse, Thomas, and Philipp Strack. 2015. “Optimal stopping with private information.” *Journal of Economic Theory* 159:702–727.

Kunstman, Jonathan W., E. Ashby, Kate Zielaskowski, and Jennifer LaCosse. 2013. “Feeling in with the outgroup: Outgroup acceptance and the internalization of the motivation to respond without prejudice.” *Journal of Personality and Social Psychology* 105 (3): 443–457. https://doi.org/10.1037/a0033082.

Kurzban, Robert, John Tooby, and Leda Cosmides. 2001. “Can race be erased? Coalitional computation and social categorization.” *Proceedings of the National Academy of Sciences* 9 (26):15387–15392. https://doi.org/10.1073/pnas.251541498.

Levine, Ross, Alexey Levkov, and Yona Rubinstein. 2008. “Racial Discrimination and Competition.” *NBER Working Paper.*

liga3-online. 2015. “Social-Media-Tabelle der 3. Liga.” Accessed July 31, 2023. https://www.liga3-online.de/social-media-tabelle-der-3-liga/.

Massey, Douglas S. 2011. “The New Immigrant Survey and Research on American Stratification.” *Social Science Research* 40 (5): 1287–1291. https://doi.org/10.1016/j.ssresearch.2011.06.001.

Maxwell, Rahsaan. 2019. “Cosmopolitan immigration attitudes in large European cities: Contextual or compositional effects?” *American Political Science Review,* 1–19. https://doi.org/10.1017/S0003055418000898.

McDonald, Jared. 2019. “Avoiding the Hypothetical: Why “Mirror Experiments” are an Essential Part of Survey Research.” *International Journal of Public Opinion Research,* https://doi.org/10.1093/ijpor/edz027.

Neumark, David. 2018. “Experimental Research on Labor Market Discrimination.” *Journal of Economic Literature* 56 (3): 799–866. Accessed May 5, 2021. https://doi.org/10. 1257/jel.20161309.

Norris, Davon, and Corey Moss-Pech. 2021. “White Men Can’t Jump, but Does It Even Matter? Exit Discrimination in the NBA.” *Social Forces,* https://doi.org/10.1093/sf/soab009.

Pager, Devah. 2016. “Are Firms That Discriminate More Likely to Go Out of Business?” *Sociological Science* 3:849–859. https://doi.org/10.15195/v3.a36.

Pardos-Prado, Sergi, and Carla Xena. 2019. “Skill Specificity and Attitudes toward Immigration.” *American Journal of Political Science* 63 (2): 286–304. https://doi.org/10.1111/ajps.12406.

Quillian, Lincoln, John J Lee, and Mariana Oliver. 2020. “Evidence from Field Experiments in Hiring Shows Substantial Additional Racial Discrimination after the Callback.” *Social Forces* 99 (2): 732–759. https://doi.org/10.1093/sf/soaa026.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. 2017. “Meta-analysis of field experiments shows no change in racial discrimination in hiring over time.” *Proceedings of the National Academy of Sciences* 114 (41): 10870–10875. https://doi.org/10.1073/pnas.1706255114.

Rivera, Lauren A. 2020. “Employer Decision Making.” *Annual Review of Sociology* 46 (1): 215–232. https://doi.org/10.1146/annurev-soc-121919-054633.

Ruedin, Didier. 2020. “Do We Need Multiple Questions to Capture Feeling Threatened by Immigrants?” *Political Research Exchange* 2 (1): 1758576. https://doi.org/10.1080/2474736X.2020.1758576.

Sailes, Gary A. 1993. “An Investigation of Campus Stereotypes: The Myth of Black Athletic Superiority and the Dumb Jock Stereotype.” *Sociology of Sport Journal* 10(1): 88–97. https://doi.org/10.1123/ssj.10.1.88.

Schildberg-Ho¨ risch, Hannah. 2018. “Are Risk Preferences Stable?” *Journal of Economic Perspectives* 32, no. 2 (May): 135–54. https://doi.org/10.1257/jep.32.2.135.

Schindler, David, and Mark Westcott. 2021. “Shocking racial attitudes: black GIs in Europe.” *The Review of Economic Studies* 88 (1): 489–520.

Seale, Darryl, and Amnon Rapoport. 1997. “Sequential decision making with relative ranks: An experimental investigation of the “secretary problem”.” *Organizational behavior and human decision processes* 69 (3): 221–236.

Weichselbaumer, Doris. 2020. “Multiple Discrimination against Female Immigrants Wearing Headscarves.” *ILR Review* 73 (3): 600–627. https://doi.org/10.1177/0019793919875707.