Racial Bias in Team Selection: Experimental Evidence from a Fictional Soccer Lineup*

Henrik Guhling† Fabian Herweg‡

August 10, 2026

We propose a controlled and incentive-compatible design to study racial bias in team-selection decisions. In an online experiment, participants select a soccer lineup from fictitious players whose objective quality is described by numerical attributes. For each player profile, we randomly assign an AI-generated portrait depicting either a Black or a White player, allowing us to isolate the effect of racial appearance. Overall, we find no systematic evidence that White player versions are selected more frequently than otherwise identical Black versions. We also compare generic-team and national-team framing, but find no systematic amplification of racial bias.

JEL classification: C91; D91; J15; J71.

Keywords: incentivized experiment; hiring and selection; racial discrimination; soccer

*We would like to thank Joshua Greubel, Maximilian Kähny, Alicia von Schenk, and participants at the University of Bayreuth graduate seminar for helpful comments and suggestions.

†University of Bayreuth, Faculty RW, Universitätsstr. 30, D-95440 Bayreuth, Germany, E-mail address: henrik.guhling@uni-bayreuth.de.

‡University of Bayreuth, Faculty RW, Universitätsstr. 30, D-95440 Bayreuth, Germany, E-mail address: fabian.herweg@uni-bayreuth.de.

1. Introduction

Racist abuse has long been a persistent feature of professional sports, particularly in the behavior of fans and spectators, both in stadiums and online. In soccer, prominent examples range from the abuse directed at John Barnes in England (BBC, 2018) to repeated attacks on Real Madrid's Vinícius Júnior, which culminated in prison sentences and stadium bans for perpetrators in Spain in 2024 (ESPN, 2024). Racist online abuse following the men's Euro 2020 final prompted police investigations and arrests in England (Sky Sports, 2021), while Germany has likewise experienced prominent recent incidents (FC Schalke 04, 2025). Such abuse is neither confined to men's soccer nor to soccer more generally: during Euro 2025, England defender Jess Carter reported racist attacks and temporarily withdrew from social media (Sky Sports, 2025). Comparable incidents have involved Formula 1 driver Lewis Hamilton (FIA, 2021), German basketball player Dennis Schröder (FIBA, 2025), and high-profile athletes in the United States (CBS News, 2017).

Whether such incidents reflect the actions of a small number of "bad actors" or instead reveal broader societal attitudes has been widely debated in public discourse. In Germany, this debate intensified ahead of the men's Euro 2024, when a survey commissioned by public-service broadcasters reported that roughly one-fifth of respondents agreed with the statement (in the following, the *national-team statement*) that they would prefer "more white players" in the German national team (infratest dimap, 2024).[1] This raises two questions. First, do such attitudes affect individual selection decisions made privately and without observation by others, even when discrimination is costly because favoring members of a particular racial group may reduce team strength, performance, and ultimately the selector's own payoff? Second, are racial considerations particularly influential in nationality-salient environments, such as the selection of national teams representing countries whose populations have historically been predominantly White?

We address these questions in a controlled and fully incentivized online experiment. Participants select a starting lineup from a roster of fictitious soccer players. Every player is characterized by two numerical attributes—offensive and defensive ability—which jointly determine an objective, position-specific measure of playing quality. Because assessing overall quality requires participants to combine these attributes using position-dependent weights, the ranking is not always immediately apparent, creating scope for the use of heuristics in selection decisions. Alongside this performance information, participants observe an AI-generated portrait of each player. For every player profile, we randomly

[1] In a similar vein, ahead of the semifinal between France and Spain at the 2026 FIFA World Cup, former Spanish prime minister Mariano Rajoy wrote of the French team: "They've won every match they've played in this World Cup and are currently ranked No. 1 in the FIFA rankings. They also have a top-level squad. That said, they don't have any French players" (The Guardian, 2026).

assign a portrait depicting either a Black or a White player, allowing us to isolate the effect of racialized appearance while holding objective quality constant. Participants' payments depend on the overall strength of the lineup they select; hence, favoring players based on race can reduce both team performance and the decision-maker's own payoff. To sharpen identification, we construct pairs of "players at the margin" whose objective qualities are very similar. In these close decisions, racial appearance may serve as a tie-breaker whenever it enters participants' preferences or heuristics. Finally, we vary the salience of national identity: in the baseline treatment, participants select a neutrally framed soccer team, whereas in the national-team treatment, they are asked to select the German national team.

With our experimental design, we address three main empirical questions. First, holding objective quality constant, are players depicted as White selected more frequently than players depicted as Black? Second, are race-dependent differences concentrated among marginal players—players of similar quality—between whom participants are likely to be close to indifferent and may use racial appearance to resolve that indifference? Third, does national-team framing amplify or attenuate racial bias in selection decisions?

Our results reveal a clear overall pattern, together with more nuanced subgroup and framing effects. First, we find no systematic evidence that players depicted as White are favored over otherwise identical players depicted as Black. Black-White differences in selection and correctness—selecting a player when selection is optimal and not selecting him otherwise—are generally small and statistically insignificant. This null result holds both in the aggregate and for choices between players at the margin, and it persists under national-team framing.

More differentiated patterns emerge in subgroup analyses. A particularly informative distinction is between participants who *reject* the *national-team statement* and those who do *not reject* it. The latter group may be more likely to hold explicit or implicit racial preferences, especially in the context of selecting a national team. Consistent with this interpretation, participants in the Rejection group perform better overall. At the same time, they exhibit a small tendency to over-select Black versions among clearly top- and bottom-ranked players, where the objectively correct choice is least ambiguous. This pattern is consistent with overcorrection or experimenter-demand effects. A separate finding concerns marginal choices under national-team framing. Here, errors tend to increase when the better player is depicted as Black and decrease when he is depicted as White. In the full sample, the clearest treatment-induced shift occurs when both marginal players are depicted as Black, and this effect is stronger among participants who do not reject the *national-team statement*.

The paper contributes in three ways. First, it provides an incentive-compatible experimental framework to test whether discriminatory behavior persists when decision-makers bear the direct cost of biased selection or employment decisions—a design that can be applied to other labor-market contexts. Second, it provides evidence on racial bias in soccer-related selection decisions under controlled incentives. Third, it examines whether national-identity framing systematically shifts behavior relative to generic team selection, using the German men's national team as the empirical setting.

The remainder of the paper is organized as follows. Section 1 concludes with a discussion of the related literature. Section 2 presents the experimental design and procedures. Section 3 derives the testable hypotheses. Section 4 reports the main empirical findings, while Section 5 presents additional exploratory analyses. Section 6 concludes.

Related Literature

The literature distinguishes between statistical discrimination, i.e., differential treatment in response to observed empirical regularities, and taste-based discrimination, i.e., differential treatment resulting from the racial preferences of the decision-maker (Lang and Kahn-Lang Spitzer, 2020). We contribute to the literature on taste-based discrimination, as our experimental design makes objective player quality observable. This eliminates the informational basis for statistical discrimination.

We mainly contribute to three strands of the literature. First, there is a sizeable literature in economics on taste-based racial discrimination in labor markets. This literature distinguishes between employer discrimination, coworker discrimination, and customer discrimination. Evidence of employer discrimination is provided by Bertrand and Mullainathan (2004), who show that resumes with White-sounding names are more likely to receive callbacks than resumes with Black-sounding names. Similarly, Giuliano et al. (2009) document racial and ethnic favoritism in hiring. Turning to coworker discrimination, they further show that employees remain with the firm longer when a larger share of their coworkers shares their ethnicity. Doleac and Stein (2013) find that, in online markets, buyers are less likely to make offers to Black sellers than to White sellers, consistent with customer-based discrimination. A survey of this literature is provided by Lang and Kahn-Lang Spitzer (2020). We contribute to this literature by providing experimental evidence on taste-based employer discrimination. However, the participants acting as team managers—and thus as employers—are recruited, at least to some degree, from the population of fans and spectators, i.e., customers, rather than from the population of professional team managers. Our findings therefore also provide evidence relevant to customer-based discrimination.

Second, we contribute to the experimental literature on racial discrimination. Studying a laboratory labor market, Wozniak and MacNeill (2020) provide evidence of taste-based discrimination by non-Black participants against Black participants. In a laboratory public-good and group-formation setting, Castillo and Petrie (2010) show that subjects rely on race to predict behavior when payoff-relevant information is absent, but that the role of race diminishes once informative signals about behavior are provided. A meta-analysis of economic experiments on discrimination is provided by Lane (2016). He documents that several studies of taste-based discrimination report null results, particularly those examining ethnic or racial discrimination. He further argues that this type of discrimination is difficult to detect in laboratory experiments. We add to this literature by providing evidence from a controlled online experiment framed as a personnel-selection problem in professional sports.

Finally, there is an extensive literature investigating discrimination in sports and related settings. Based on historical data from the NBA, Zhang (2017) reports that players receive more playing time under coaches of the same race. Regarding soccer, Chu et al. (2014) document differential treatment of White and non-White players in outcomes such as team selection and disciplinary decisions in the English Premier League. For German soccer, specifically the first and second Bundesliga, Nobis and Lazaridou (2023) document evidence consistent with *racist stacking*: White players are disproportionately assigned to positions associated with leadership and oversight, whereas Black players are overrepresented in positions stereotypically associated with speed and aggressiveness. Using a field experiment on invitations to soccer tryouts in Scandinavian countries, Storm et al. (2023) find evidence of discrimination against applicants with foreign-sounding names in Sweden, but not in the other countries studied. More closely related to traditional labor-market studies, Szymanski (2000) finds evidence of a wage gap for Black players in English professional soccer during the 1980s, whereas Liang et al. (2023) report no racial wage gap in more recent NBA data. A related literature uses data from fantasy-sports platforms, which constitute virtual player-selection settings. Using data from the Fantasy Premier League, Bryson and Chevalier (2015) find no evidence of racial discrimination in player-selection decisions. Similarly, Kotrba and Dwyer (2023) report no discriminatory selection patterns in Premier League-related fantasy choices among a sample of Czech gamers. In contrast, Kotrba (2021) documents a preference for domestic players in Czech fantasy-soccer settings. We contribute to this literature by introducing a controlled, incentive-compatible experiment that resembles a fantasy-sports selection task and links participants' payments directly to the strength of the selected team. This design allows us to test whether racial cues affect selection when discrimination is privately costly and objective information about player quality is available.[2]

[2]There is also evidence on customer-based discrimination—that is, discrimination by fans—in sports. Meier and Leinwather (2013) analyze television audiences for matches of the German national team

2. Experimental Design and Procedures

2.1. Experimental Design

We employ a between-subjects design to address the research questions outlined above. Each participant is asked to select the best 11 players for a starting lineup from a roster of 22 fictitious soccer players. The task is decomposed into four subtasks: selecting one of two `goalkeepers`, three of six `defenders`, four of eight `midfielders`, and three of six `attackers`. Each player is described by two numerical attributes, an *offensive* value and a *defensive* value. The quality of player j—his performance index $\mathcal{P}_j$—is defined as a position-specific weighted average of these two values:

$$\mathcal{P}_j = \alpha^{pos} \times \text{defensive} + (1 - \alpha^{pos}) \times \text{offensive},$$

where $\alpha^{pos} = 1$ for goalkeepers, $\alpha^{pos} = 0.72$ for defenders, $\alpha^{pos} = 0.48$ for midfielders, and $\alpha^{pos} = 0.28$ for attackers. Team strength is defined as the unweighted average of the performance indices of the 11 selected players.

(a) Attacker depicted as Black.

(b) The same attacker depicted as White.

Figure 1: Example of the two racial-appearance versions of the same player profile.

In addition to the numerical attributes, each player is visually represented by an AI-generated facial image. Our key manipulation is a randomized visual cue of race: for each player, we randomly assign an AI-generated portrait signaling either a Black or a White appearance. For every player profile, we generated two portraits, one Black and one White

and find little evidence of discrimination. Maennig and Mueller (2022) report evidence of both consumer- and employer-side discrimination in Major League Baseball. Investigating the five major European soccer leagues, Quansah et al. (2024) find that stadium attendance in Italy is linked to the racial composition of the teams.

(see Figure 1).[3] This two-version structure allows us to compare selection frequencies for the same player profile across racial appearance.

To examine how national-identity framing affects selection decisions, we implement two treatments. In the baseline treatment, participants select the starting lineup for a generic soccer team, whereas in the national-team treatment they select the starting lineup for the German men's national team. Specifically, participants in the national-team treatment are told that they are the head coach of the German men's national team and must select the starting lineup for the upcoming matches.[4] In addition, players in this treatment wear jerseys displaying the colors of the German national flag, whereas players in the baseline treatment wear neutral jerseys (see Figure 2).

(a) A defender under generic-team framing. (b) Same defender under national-team framing.

Figure 2: Example of the generic-team and national-team framing.

Each participant selects 11 players from an identical roster of 22 players with respect to objective quality. Choices are incentivized through the simulation of 10 matches at the end of the experiment, with each win yielding an additional payment of £0.50. In each match, the participant's selected team competes against a benchmark team constructed ex ante from the same roster. A match is coded as a win if the participant's team has a higher team strength than the benchmark team and as a loss otherwise (draws cannot occur by construction). Thus, participants are financially incentivized to select the objectively strongest possible lineup.

The task is designed such that identifying the optimal lineup is difficult without careful calculation and that reliance on heuristics is therefore likely to generate some mistakes,

[3]Perceived race was validated by independent raters in a separate classification task (not by the main-study participants). Furthermore, we took particular care to ensure that the two player versions resembled each other as closely as possible in terms of facial features without appearing unnatural. An overview of all player versions is provided in Appendix A.

[4]An English translation of the instructions is provided in Appendix A.2.

particularly for player pairs at the selection margin. Specifically, for defenders, midfielders, and attackers, we construct two near-substitute candidates whose performance indices are deliberately very similar. In these close comparisons, participants who consciously or unconsciously favor the White player may rely on a heuristic that provides a justification for selecting him, even when the Black player is of slightly higher objective quality. Choices between players at the margin can therefore reveal implicit racial preferences.

2.2. Procedures

The experiment was preregistered on AsPredicted (available under https://aspredicted.org/9rx8cp.pdf).[5] It was computerized using o-Tree (Chen et al., 2016) and conducted online via Prolific in December 2025 and January 2026. In total, we recruited 1,204 German-speaking participants residing in Germany. The final sample comprises 1,200 participants (421 female, 768 male, and 11 non-binary), with 600 assigned to each of the generic-team and national-team treatments.

The average payment received by participants was £4.66, including a fixed participation fee of £2.50. The median completion time was 9.6 minutes. Before making any choices, participants received detailed instructions on the task, the construction of the performance indices, the definition of team strength, and the payment rule. They then answered comprehension questions to verify their understanding of the task and procedure. The four participants excluded from the final sample either answered a comprehension question incorrectly twice or did not complete the experiment.[6]

The selection task proceeded sequentially. Participants first selected the goalkeeper, followed by the defenders, midfielders, and attackers. After completing these decisions, they were shown the full provisional lineup and could revise their selections before submitting it.

After the main task, participants completed a questionnaire. We asked about their familiarity with computer games involving similar tasks, their self-assessed mental-arithmetic skills, their interest in soccer and the German men's national team, and whether they actively played or coached soccer. We also elicited the German political party to which they felt closest and recorded their agreement with one version of the *national-team statement*: "The share of light-skinned players in the German men's national soccer team should be increased." Finally, we collected sociodemographic information on gender, age, origin, and migration background.

[5]IRB approval was granted by the German Association for Experimental Economic Research e.V. (GfeW).

[6]The exclusion rule and the target of 600 participants per treatment were preregistered. Participants excluded from the final sample could still receive the fixed fee of £2.50 if they completed the experiment.

3. Hypotheses

With our design, we aim to test whether subtle racist tendencies are reflected in behavior. We are primarily interested in whether White players are selected more often, but also whether any such tendencies—if present—are amplified in near-indifferent choices ("players at the margin") and whether behavior differs between the generic-team treatment and the national-team treatment.

If subtle racist tendencies exist, we expect that, holding quality constant, the White version of a player is chosen more frequently.

Hypothesis 1 (Racial Bias Effect)**.** *Conditional on objective quality, players depicted as White are more likely to be selected than otherwise identical players depicted as Black.*

Since the top- and bottom-ranked players are relatively straightforward to identify, we expect racial tendencies to matter little at the extremes of the quality distribution. The most informative tests therefore concern close decisions involving players at the margin. We construct these marginal matchups as comparisons between two players with nearly identical performance indices, such as the third- and fourth-best defender—the last player who should be selected and the first player who should not be selected. Choices at the margin thus approximate a binary decision in which racial appearance may serve as a tie-breaker. If subtle racial bias is present, participants should be less likely to select the higher-quality player when he is depicted as Black rather than White. The largest difference should arise between mixed-race pairings in which the better player is White and the worse player is Black, and pairings in which the better player is Black and the worse player is White.[7]

Hypothesis 2 (Marginal Player Effect)**.** *For choices between two players at the margin, the probability of selecting the higher-quality player is highest when he is depicted as White and the lower-quality player as Black, and lowest when the racial configuration is reversed. The corresponding probabilities for White–White and Black–Black pairs lie between these two cases.*

Finally, we examine whether behavior differs between the generic-team and national-team treatments. If subtle racial bias leads individuals to associate the German men's national team more strongly with White players, the effects described in the two preceding hypotheses should be more pronounced under national-team framing.

Hypothesis 3 (Treatment Effect)**.** *The racial bias effect and the marginal player effect are stronger under the national-team treatment than under the generic-team treatment.*

[7] Selecting one marginal player rather than the other has only a small effect on team strength. Nevertheless, this difference may affect the number of wins in the ten simulated matches.

4. Results

This section presents the experimental results for our three hypotheses. For the analyses of Hypotheses 1 and 2, we pool the data from the two treatments, whereas Hypothesis 3 concerns differences between the generic-team and national-team treatments. Before turning to the hypothesis tests, we note that respondents who identify with parties on the left of the political spectrum are overrepresented in our sample (see Appendix Table 4); we discuss the implications for external validity in Section 5. Appendix Table 5 further documents small random imbalances in some baseline characteristics between the two treatments. As a robustness check, we therefore re-estimate the main specifications controlling for these characteristics.

4.1. Hypothesis 1: Selection of Black vs. White Player Versions

To test Hypothesis 1, we proceed in two steps. First, we descriptively compare the selection rates of the Black and White versions of each player. Second, we conduct an aggregate analysis by grouping players into three categories. The clearly best players at each position are assigned to the `in` category, whereas the clearly worst players are assigned to the `out` category. The two players at the selection margin—the lowest-quality player who should be selected and the highest-quality player who should not be selected—are assigned to the `marginal` category. For example, among defenders, the two players ranked first and second in objective quality belong to the `in` category, the players ranked third and fourth belong to the `marginal` category, and the players ranked fifth and sixth belong to the `out` category.[8] Based on this classification, we compare how often players of each category are assigned correctly, that is, selected when selection is optimal and not selected when selection is suboptimal. While this measure contains the same information as the selection rate for `in` and `out` players, it is more informative for `marginal` players because this category includes some players who should be selected and some who should not. We then investigate whether these correctness rates differ between the Black and White versions of otherwise identical players.[9]

Figure 3 provides descriptive evidence for Hypothesis 1 and confirms that the task generated the intended differences in decision difficulty. Across positions, selection rates closely follow the objective ranking implied by the performance index. Clearly dominant players are selected at very high rates, whereas clearly dominated players are rarely selected. Importantly, mistakes are concentrated where the task is designed to be difficult. For defenders, midfielders, and attackers, two players exhibit intermediate selection rates,

[8]Because the choice between the two goalkeepers is particularly clear-cut, we classify the better goalkeeper as `in` and the worse goalkeeper as `out`; hence, there is no `marginal` player for this position.

[9]A detailed description of the empirical strategy used to test all hypotheses is provided in Appendix B.

consistent with our design objective of creating *players at the margin* for whom identifying the better player requires more careful calculation and errors are therefore more likely.[10]

Turning to racial differences in selection, Figure 3 shows no systematic tendency for White versions to be selected more frequently than Black versions at the player level. The vast majority of Black–White gaps are small and statistically indistinguishable from zero; the corresponding estimates and p-values are reported in Appendix Table 6. If anything, the few statistically significant differences tend to point toward slightly higher selection rates for the Black version, although they arise for only a small number of players.

To summarize the evidence more compactly, Table 1 aggregates outcomes by player category (`in`, `marginal`, and `out`) and reports correctness rates. Consistent with Figure 3, we find no indication that White versions are favored on average. If anything, Black versions of players in the `in` category are assigned correctly somewhat more often—that is, they are selected when they should be selected—by about 1.0–1.2 percentage points. This difference becomes weakly statistically significant at the 10% level when controlling for participant-level characteristics. This suggests that, in some cases, White player versions may be selected somewhat less frequently. We return to this pattern in Section 5.

Taken together, Figure 3 and Table 1 do not support Hypothesis 1. In particular, the 95% confidence intervals allow us to rule out any systematic preference for White player versions exceeding 0.4 percentage points across the `in` and `out` categories. Because aggregate correctness for `marginal` players inherently mixes the decisions to select the better players and to omit the worse ones, a single pooled bound cannot cleanly isolate a directional preference. We therefore evaluate these specific binary choices in detail below when testing Hypothesis 2.

Table 1: Decision Accuracy by Player Category and Racial Appearance

	(1)	(2)
In	0.883***	0.790***
	(0.005)	(0.039)
Marginal	0.493***	0.400***
	(0.011)	(0.040)
Out	0.899***	0.806***
	(0.005)	(0.039)

Continued on next page

[10] In the preregistration, we anticipated four marginal players in midfield. Based on the realized selection rates in Figure 3, however, midfield exhibits two clear "knife-edge" profiles, while the remaining players are substantially more (or less) likely to be selected. We therefore classify only these two midfield players as marginal in the analysis.

	(1)	(2)
In × Black	0.010	0.012*
	(0.006)	(0.006)
Marginal × Black	0.006	0.006
	(0.012)	(0.012)
Out × Black	-0.008	-0.007
	(0.006)	(0.006)
Controls	No	Yes
Observations	26,400	26,400
R^2	0.823	0.826

Notes: The dependent variable is whether the player's classification was correct. Model (1) includes no participant-level controls. Model (2) adds participant-level controls. The full regression including all estimates can be seen in Appendix D, Table 7. CR2 cluster-robust standard errors at the participant level are reported in parentheses.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Result 1. *Across players and player categories, we find no systematic evidence that White versions are more likely to be selected or correctly assigned. If anything, there are slight tendencies toward higher selections for Black versions among* `in` *players.*

4.2. Hypothesis 2: Racial Cues in Marginal Choices

We next turn to Hypothesis 2 and focus on the players at the `margin`. The marginal set comprises, for each of defense and attack, the third- and fourth-best players, and for midfield the fourth- and fifth-best player. We therefore classify two players per position as marginal, yielding a set of binary comparisons in which participants must decide between two near-substitutes at the selection cutoff.

More specifically, we are primarily interested in cases in which exactly one of the two *players at the margin* was selected. We therefore examine the correctness rate, defined as the share of these cases in which the better, and thus correct, player was chosen, and compare it across the four possible race configurations (`Black/Black`, `Black/White`, `White/Black`, and `White/White`).[11] Hypothesis 2 predicts that this correctness rate should be highest in the `White/Black` configuration and lowest in the `Black/White` con-

[11] Considering only cases in which exactly one of the players was selected has the advantage of approximating the binary choice as closely as possible. In most cases, indeed, exactly one of the two players was selected (defense: 91%; midfield: 74%; attack: 93%). As a robustness check, however, we also consider the full set of observations and report how often the better player was selected in these configurations. Results can be seen in Appendix D, Table 9.

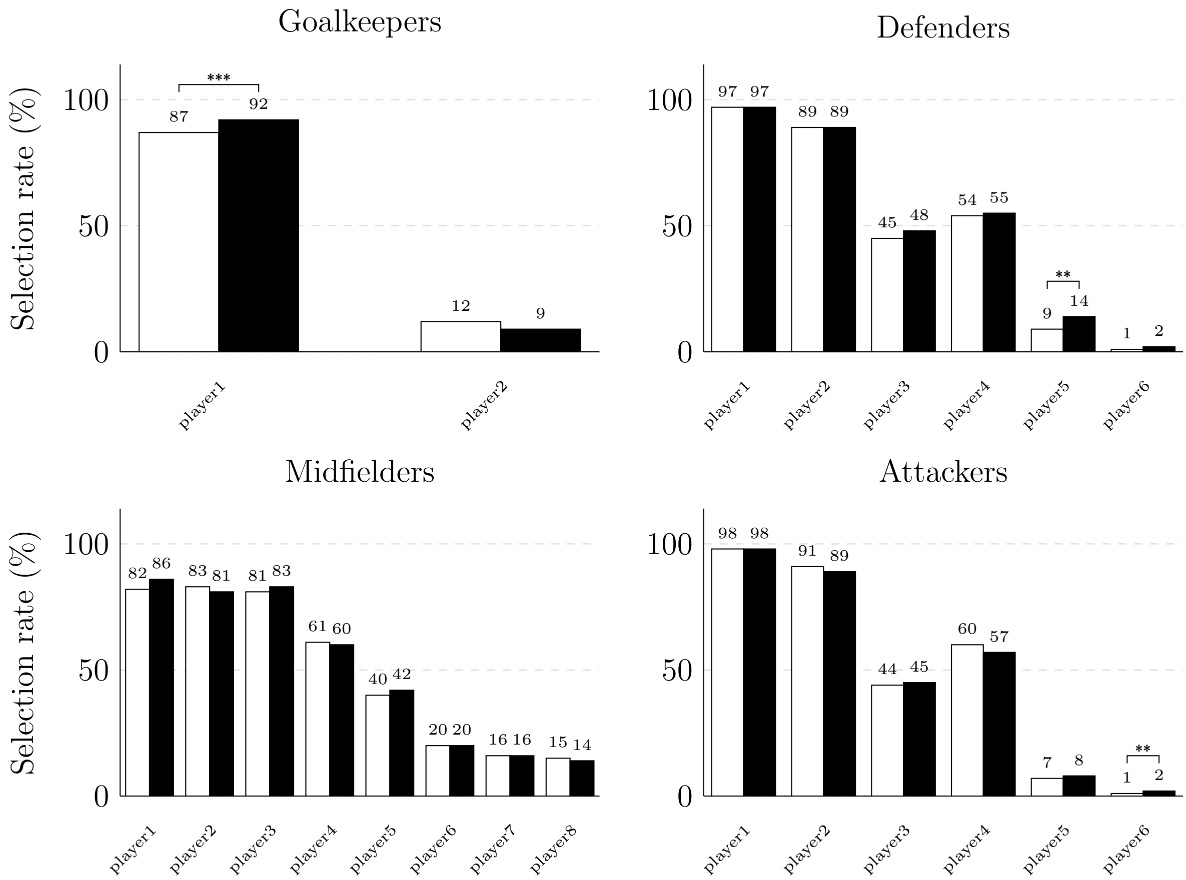


Figure 3: Selection rates by position with players ranked by objective quality within each position. White bars indicate players depicted as White, and black bars indicate players depicted as Black. Estimates and p-values are based on a linear probability model without intercept and with player fixed effects.

figuration.

Figure 4 compares these correctness rates across the four possible race configurations. We find no substantial differences across the configurations. Although the correctness rate is indeed lowest in the `Black/White` configuration, and the rates tend to be somewhat higher when the worse player is displayed as Black, the predicted overall ranking does not emerge. Moreover, all correctness rates lie between 48% and 51% and are not statistically distinguishable from one another. We therefore find no evidence in support of Hypothesis 2. In particular, based on the 95% confidence intervals, we can rule out that, compared to a White/White baseline, the Black/White configuration reduces correct selection by more than 5.4 percentage points, or that the White/Black configuration increases it by more than 6.1 percentage points. As expected, these bounds are somewhat wider than in the analysis for Hypothesis 1 because restricting the focus to binary marginal choices

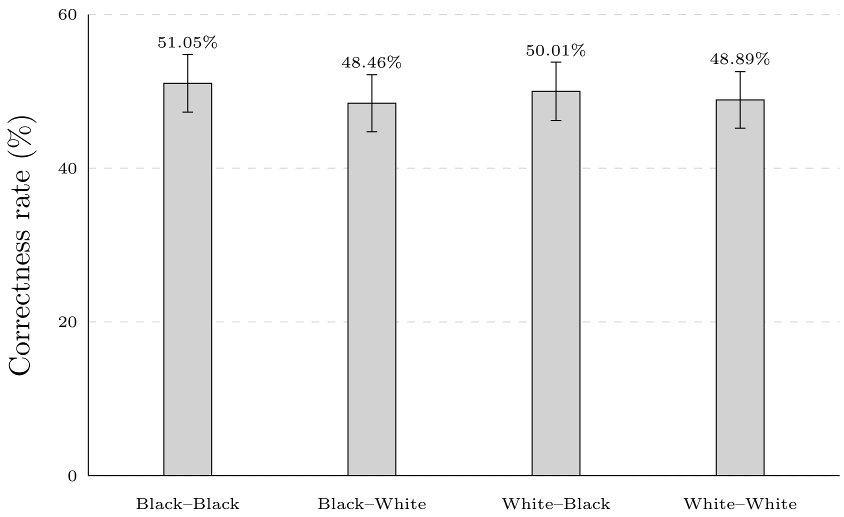


Figure 4: Correctness rates by race configuration. Error bars indicate 95% confidence intervals. Estimates are based on a linear probability model. The full model can be seen in Appendix D, Table 8.

reduces the sample size.

Result 2. *Overall, we find no evidence that White player versions have a selection advantage among the players at the margin. The higher-quality player is selected somewhat more frequently when the lower-quality player is depicted as Black, but the differences across racial-appearance configurations are not statistically significant.*

4.3. Hypothesis 3: Effects of National-Team Framing

To examine Hypothesis 3, we analyze whether the effects considered in the two preceding sections differ between the two treatment groups. We first assess whether the Black–White differences in correctness rates for the three player categories, i.e., `in`, `marginal`, and `out`, vary across treatments. We then examine whether the two treatment groups differ in the correctness rates observed in the binary comparisons of the players at the `margin`.

Figure 5 displays the Black–White differences in correctness rates across the different player categories and treatment conditions, together with 95% confidence intervals. For the `in` and `out` categories, the pattern is similar across treatment conditions and mirrors that observed in the full sample. White player versions are selected slightly less often, although the differences are not statistically significant. For the `marginal` category, however, the patterns differ across treatments. In the generic-team treatment, the Black–White difference is positive (estimate: 0.024, $p = 0.153$), indicating that Black player versions are classified correctly somewhat more often, although again not significantly. In the national-team treatment, by contrast, the difference is negative (estimate:

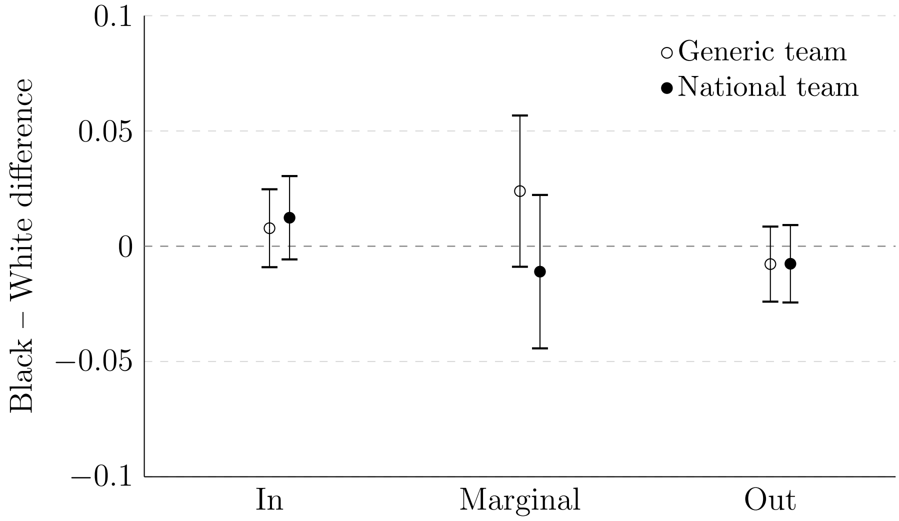


Figure 5: Black–White difference in correctness rates by player category and treatment. Points indicate estimated differences; error bars indicate 95% confidence intervals. Estimates are based on a linear probability model. The full model can be found in Appendix D, Table 10.

-0.011, $p = 0.516$), implying that White player versions are classified correctly slightly more often. Although the difference between these two Black–White effects is not statistically significant ($p = 0.143$), it is nevertheless somewhat more pronounced than the corresponding differences for the other player categories.

Table 2: Correctness Rates by Race Configuration and Treatment

	(1)	(2)
Black–Black	0.558***	0.258
	(0.027)	(0.150)
Black–White	0.493***	0.202
	(0.026)	(0.148)
White–Black	0.473***	0.178
	(0.028)	(0.150)
White–White	0.479***	0.193
	(0.026)	(0.149)
Black–Black × National Team	-0.091**	-0.090**
	(0.038)	(0.038)
Black–White × National Team	-0.018	-0.020
	(0.038)	(0.037)
White–Black × National Team	0.051	0.044

Continued on next page

	(1)	(2)
	(0.039)	(0.039)
White–White × National Team	0.020	-0.002
	(0.037)	(0.037)
Controls	No	Yes
Observations	3,100	3,100
R^2	0.497	0.514

Notes: The dependent variable indicates whether the better player in each configuration was selected. Model (1) includes no participant-level controls. Model (2) includes all participant-level controls. The full regression, including all estimates, can be found in Appendix D, Table 11. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Examining the racial-appearance configurations of the *players at the margin* across the two treatments provides further insight into the treatment-related pattern observed for the `marginal` category. Table 2 reports correctness rates for the four configurations and the corresponding treatment differences. The decline in correctness when marginal players are depicted as Black is driven primarily by the `Black/Black` configuration. Under generic-team framing, the probability of selecting the higher-quality player is highest for this configuration. The difference relative to `Black/White` is significant at the 10% level, while the differences relative to `White/Black` and `White/White` are significant at the 5% level. Under national-team framing, however, the correctness rate for `Black/Black` decreases by 9 percentage points ($p < 0.05$). In this treatment, correctness is highest for the `White/Black` configuration, although none of the corresponding differences across configurations is statistically significant.

Overall, the results provide only limited support for Hypothesis 3. In particular, we find no treatment differences for players in the `in` and `out` categories. Across these categories, the 95% confidence intervals rule out framing-induced increases of more than 0.9 percentage points in the selection of White player versions. For the players at the `margin`, neither treatment condition exhibits a consistent ordering across race configurations. Thus, we find no robust evidence in line with Hypothesis 2. On the other hand, although the point estimates for the Black/White and White/Black configurations are not statistically significant, they point in the predicted direction, and the 95% confidence intervals still leave room for modest effects. Specifically, we can only rule out treatment-induced decreases exceeding 9.2 percentage points for the Black/White configuration and treatment-induced increases exceeding 12.7 percentage points for the White/Black configuration. Lastly, we observe a significant treatment-induced drop in correctness for the Black/Black configuration, suggesting a general loss of accuracy in an all-Black context rather than a directed preference for White players.

Result 3. *We find no evidence that White player versions are systematically selected more frequently under national-team framing. The results nevertheless provide tentative evidence that national-team framing may affect the evaluation of Black players among the players at the margin.*

5. Exploratory Analysis

In this section, we examine two suggestive patterns in greater detail. First, White player versions are selected somewhat less frequently among players in the `in` and `out` categories. Second, among the players at the `margin`, we observe a significant treatment difference for the `Black/Black` configuration.

Racial attitudes may differ systematically across segments of society, and participants who identify with parties on the left of the political spectrum are overrepresented in our sample (see Appendix C, Table 4). It is therefore informative to examine whether these patterns differ across participant groups. Because responses to our version of the *national-team statement* may capture racial preferences more directly than self-reported party affiliation, we divide the sample according to participants' responses to this item.

Figure 6 reports the distribution of responses to the statement, "The share of light-skinned players in the German men's national soccer team should be increased." Approximately 60% of participants reject the statement, whereas a substantial minority of about 40% do not. Table 12 in Appendix D further shows that participants who do not reject the statement perform somewhat worse on average than those who reject it. We therefore divide the sample into a *Rejection* group and a *Non-Rejection* group.

Figure 7a reports differences in correctness rates—selecting players who should be selected and not selecting those who should not be selected—between the *Rejection* and *Non-Rejection* groups, distinguished by player category and racial appearance. The performance gap between the two groups is driven primarily by differences in the cognitively easier `in` and `out` decisions and is largely independent of racial appearance. These differences translate into an average bonus payment that is 18% lower for participants in the *Non-Rejection* group. By contrast, the performance gap does not arise for the more error-prone `marginal` players.

Figure 7b reports the Black–White differences in correctness rates for the *Rejection* and *Non-Rejection* groups. Among participants in the *Rejection* group, Black player versions are selected significantly more often than White player versions in both the `in` and `out` categories. This increases correctness for `in` players but reduces it for `out` players. By

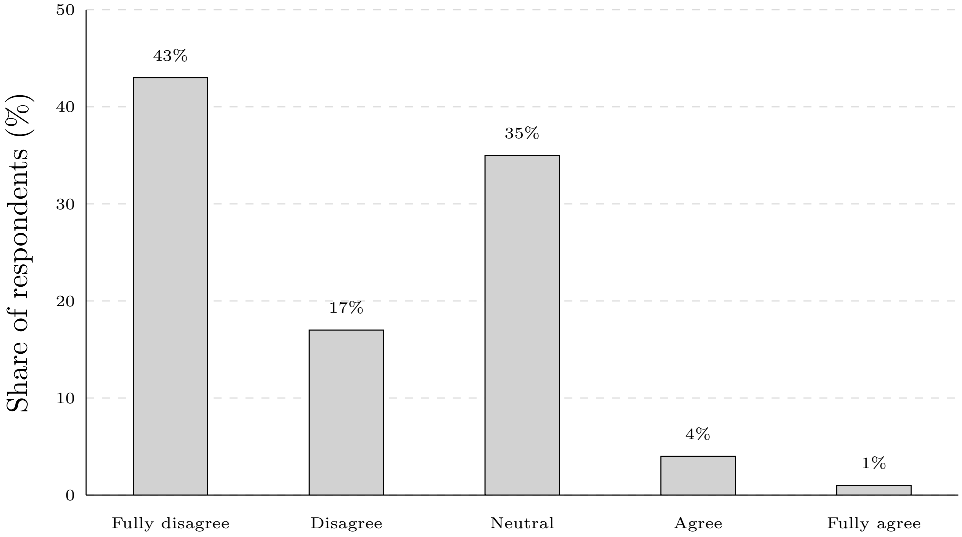
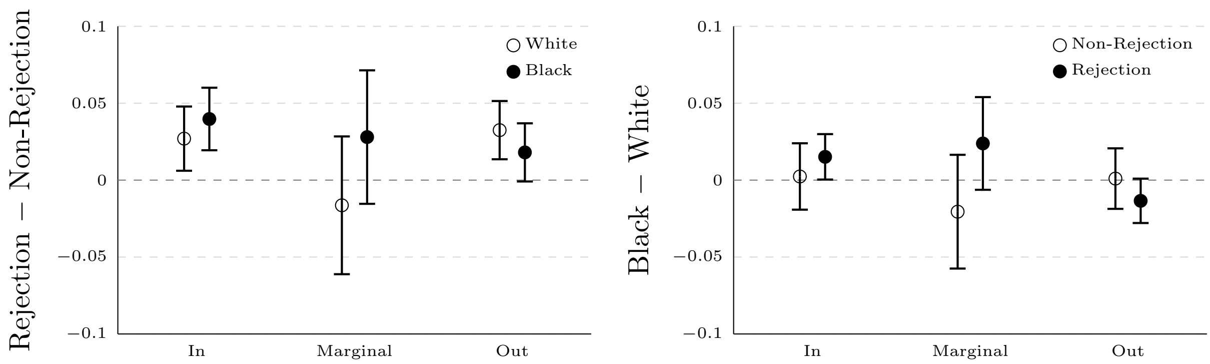


Figure 6: Distribution of responses to the statement: "The share of light-skinned players in the German men's national soccer team should be increased."



(a) Rejection versus Non-Rejection

(b) Black versus White

Figure 7: Differences in correctness rates by player category, racial appearance, and attitude group. Panel A reports Rejection–Non-Rejection differences in correctness rates separately for players depicted as White and Black. Panel B reports Black–White differences in correctness rates separately for the *Non-Rejection* and *Rejection* groups. Error bars indicate 95% confidence intervals.

contrast, no statistically significant Black–White differences emerge in the *Non-Rejection* group. The full-sample pattern whereby White player versions are selected somewhat less often in the `in` category therefore appears to be driven primarily by participants in the *Rejection* group. For the players at the `margin`, the two groups display opposing tendencies. Correct classification favors Black player versions in the *Rejection* group and White player versions in the *Non-Rejection* group, although neither difference is statistically significant.

Given these opposing tendencies for the players at the `margin`, it is useful to examine whether they are associated with particular racial-appearance configurations. Figure 8 compares correctness rates between the *Rejection* and *Non-Rejection* groups across the four configurations. Overall, the differences are small and none is statistically significant, consistent with the corresponding full-sample analysis. The most pronounced difference occurs in the `Black/Black` configuration, for which the *Rejection* group performs somewhat better. This difference may partly explain the opposing tendencies described above, but the evidence remains inconclusive. Thus, when pooling the two treatment conditions, no clear pattern emerges for the players at the `margin`.

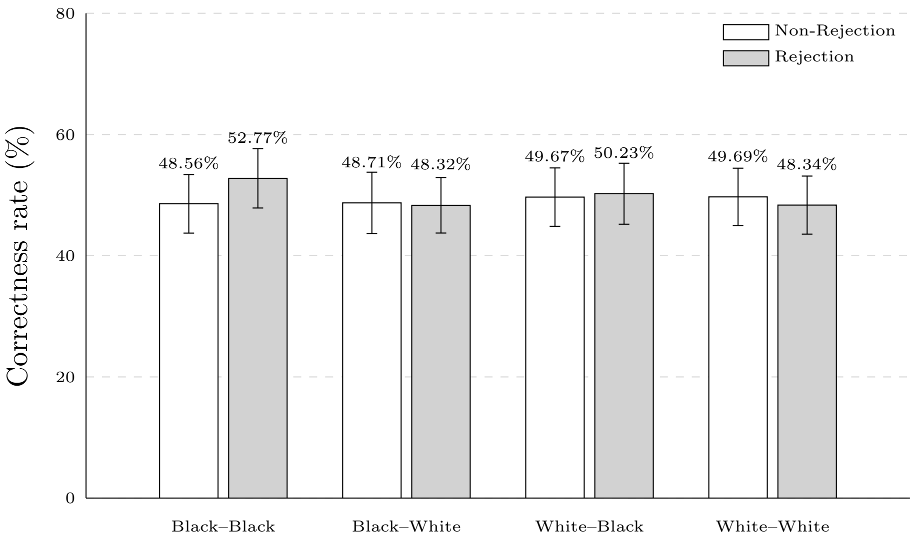
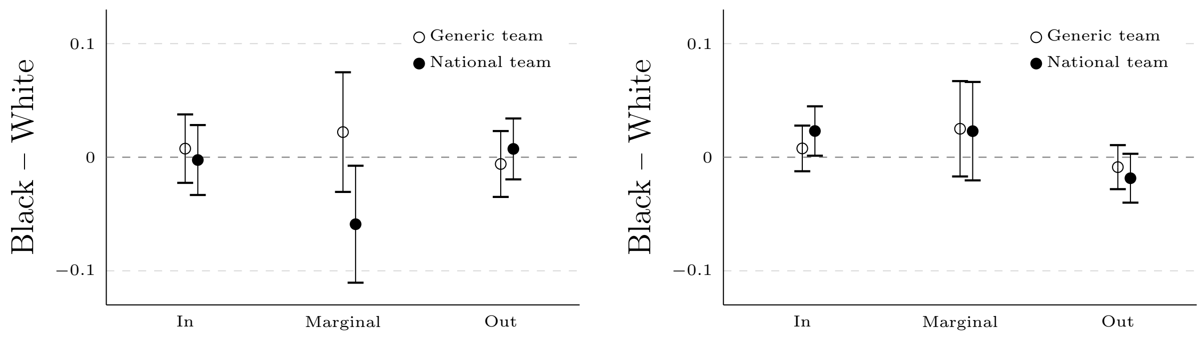


Figure 8: Selection rates by race configuration and attitude group. Error bars indicate 95% confidence intervals. Estimates are based on a linear probability model.

Therefore, Figure 9 compares the Black–White differences in correctness rates across treatment conditions separately for the *Rejection* and *Non-Rejection* groups. For the `in` and `out` categories, no significant treatment differences emerge in the *Non-Rejection* group. The same is broadly true for the *Rejection* group, although the tendency to select Black player versions more often appears to be more pronounced in the national-team treatment.

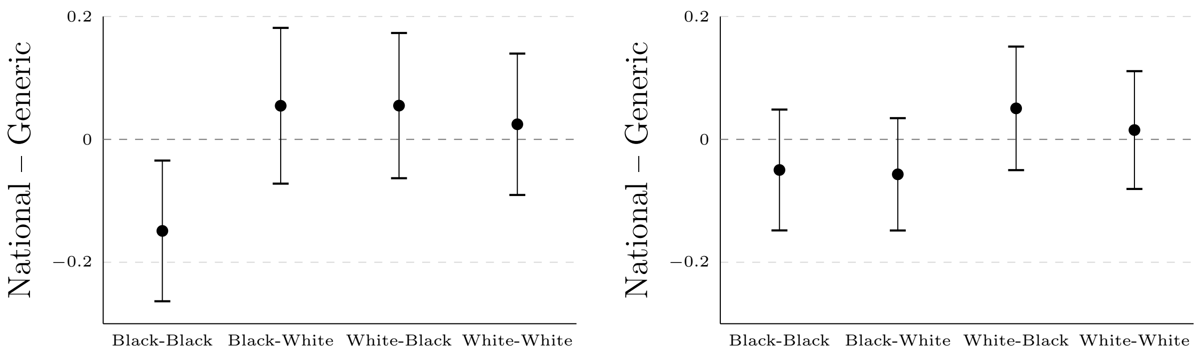
For the `marginal` category, the *Rejection* group again exhibits little variation across treatments, with Black player versions being classified correctly somewhat more often in both conditions. For the *Non-Rejection* group, however, the opposing pattern becomes more pronounced. Under generic-team framing, Black player versions are classified correctly more often, whereas under national-team framing, White player versions are classified correctly significantly more often. Although the difference between these treatment-specific Black–White gaps is not statistically significant, the pattern suggests that particular racial-appearance configurations may respond differently to national-team framing.



(a) Non-Rejection (b) Rejection

Figure 9: Black–White differences in correctness rates by treatment, attitude group, and player category. Points indicate estimated differences, and error bars indicate 95% confidence intervals.

Turning to treatment differences across racial-appearance configurations, Figure 10 shows that none of the differences in the *Rejection* group is statistically significant. Nevertheless, relative to the generic-team treatment, correctness rates under national-team framing are lower for the `Black/Black` and `Black/White` configurations, but higher for the `White/Black` and `White/White` configurations. The clearest pattern emerges in the *Non-Rejection* group. Here, the opposing effects observed in the full sample and in the preceding analysis of the `marginal` category are driven primarily by the `Black/Black` configuration. The corresponding treatment difference is statistically significant and, at approximately 15 percentage points, somewhat larger than in the full sample. This suggests that the full-sample pattern is driven mainly by participants in the *Non-Rejection* group.



(a) Non-Rejection (b) Rejection

Figure 10: Treatment differences in correctness rates by racial-appearance configuration and attitude group. Points indicate estimated differences, and error bars indicate 95% confidence intervals.

In summary, participants who reject the *national-team statement* tend to select Black player versions somewhat more frequently, particularly in cognitively easier decisions involving `in` and `out` players. This tendency is reflected to some extent in the Black–White difference for the `in` category in the full sample, but is otherwise less pronounced. Participants who do not reject the *national-team statement* generally perform somewhat worse, mainly because of lower correctness rates for `in` and `out` players. Within these categories, however, correctness does not differ between Black and White player versions. For the *players at the margin*, a different pattern emerges. Under national-team framing, Black player versions are classified incorrectly significantly more often. This result is driven primarily by the `Black/Black` configuration, in which the higher-quality player is selected substantially less often.

6. Conclusion

We study racial bias in a controlled and fully incentivized online experiment in which participants select a soccer lineup from fictitious players whose objective quality is observable. Players are randomly depicted as Black or White, and we compare a generic-team treatment with a national-team treatment. Overall, we find no systematic evidence that White player versions are favored over otherwise identical Black player versions, neither in the aggregate nor among players at the selection margin.

This null result is consistent with the conclusion of Lane (2016) that racial discrimination is difficult to detect in laboratory experiments, and suggests that the same difficulty extends to online experiments. Participants may infer the purpose of the study and choose what they regard as socially appropriate or what they believe the researchers expect. Such demand effects may be particularly strong when participants perceive researchers as holding predominantly left-leaning views. Our exploratory findings point to a possible way forward. Recruiting a politically more heterogeneous sample, including more participants who are less strongly aligned with left-wing positions, and combining this with a national-identity-salient decision environment may make discriminatory preferences easier to detect. This interpretation remains tentative, but it suggests that both sample composition and contextual salience are important for the experimental study of racial discrimination.

References

BBC, “Invisible Banana Skins Thrown at Black People Every Day, Says John Barnes,” https://www.bbc.com/sport/football/46504433 2018. Accessed July 30, 2026.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bryson, Alex and Arnaud Chevalier, "Is There a Taste for Racial Discrimination amongst Employers?," *Labour Economics*, 2015, *34*, 51–63.

Castillo, Marco and Ragan Petrie, "Discrimination in the Lab: Does Information Trump Appearance?," *Games and Economic Behavior*, 2010, *68* (1), 50–59.

CBS News, "Racial Slur Spray-Painted on Gate outside LeBron James' Home, Police Say," https://www.cbsnews.com/news/lebron-james-home-racial-slur-graffiti-brentwood-california/ 2017. Accessed February 23, 2026.

Chen, Daniel L., Martin Schonger, and Chris Wickens, "oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Chu, Jeffrey, Saralees Nadarajah, Emmanuel Afuecheta, Stephen Chan, and Ying Xu, "A Statistical Study of Racism in English Football," *Quality & Quantity*, 2014, *48* (5), 2915–2937.

Doleac, Jennifer L. and Luke C. D. Stein, "The Visible Hand: Race and Online Market Outcomes," *Economic Journal*, 2013, *123* (572), F469–F492.

ESPN, "Valencia Fans Jailed for Real Madrid's Vinícius Racial Abuse," https://www.espn.com/soccer/story/_/id/40318791/valencia-fans-jailed-real-madrid-vinicius-racial-abuse 2024. Accessed February 23, 2026.

FC Schalke 04, "Schalke 04 Strongly Condemns Racist Abuse Directed at Christopher Antwi-Adjei," https://schalke04.de/en/dfb-cup/s04-condemns-racial-abuse-antwi-adjei/ 2025. Accessed February 23, 2026.

FIA, "F1 – Joint Statement from Formula 1, the FIA and the Mercedes-AMG Petronas F1 Team," https://www.fia.com/news/f1-joint-statement-formula-1-fia-and-mercedes-amg-petronas-f1-team 2021. Accessed February 23, 2026.

FIBA, "FIBA Statement on Incident during FIBA EuroBasket 2025 Game between Germany and Lithuania," https://www.fiba.basketball/en/events/fiba-eurobasket-2025/news/fiba-statement-on-incident-during-fiba-eurobasket-2025-game-between-germany-lithuania 2025. Accessed February 23, 2026.

Giuliano, Laura, David I. Levine, and Jonathan Leonard, "Manager Race and the Race of New Hires," *Journal of Labor Economics*, 2009, *27* (4), 589–631.

infratest dimap, "Deutsche Fußballmannschaft 2024: Mehrheit findet multikulturelles Team gut," https://www.infratest-dimap.de/umfragen-analysen/bundesweit/umfragen/aktuell/deutsche-fussballmannschaft-2024-mehrheit-findet-multikulturelles-team-gut/ 2024. Accessed February 23, 2026.

Kotrba, Vojtěch, "Nationality-Based Prejudice in Sports: Fan Demand for Foreign Athletes in the Czech Republic," *International Journal of Sport Finance*, 2021, *16* (1), 3–15.

_ **and Brendan Dwyer**, "Examining Racial Discrimination in Fantasy Team Selection: The Case of Czech Soccer Fans," *Sport, Business and Management: An International Journal*, 2023, *13* (2), 248–267.

Lane, Tom, "Discrimination in the Laboratory: A Meta-Analysis of Economics Experiments," *European Economic Review*, 2016, *90*, 375–402.

Lang, Kevin and Ariella Kahn-Lang Spitzer, "Race Discrimination: An Economic Perspective," *Journal of Economic Perspectives*, 2020, *34* (2), 68–89.

Liang, JiaYou, ShuaiJie Zhao, and HaoYuan Zhu, "Is There Salary Discrimination by Race and Nationality in the NBA? A New Approach," in "International Conference on Economic Management and Green Development" Springer 2023, pp. 391–399.

Maennig, Wolfgang and Steffen Q. Mueller, "Consumer and Employer Discrimination in Professional Sports Markets: New Evidence from Major League Baseball," *International Journal of Sport Finance*, 2022, *17* (4), 230–244.

Meier, Henk Erik and Marcel Leinwather, "Finally a 'Taste for Diversity'? National Identity, Consumer Discrimination, and the Multi-Ethnic German National Football Team," *European Sociological Review*, 2013, *29* (6), 1201–1213.

Nobis, Tina and Felicia Lazaridou, "Racist Stacking in Professional Soccer in Germany," *International Review for the Sociology of Sport*, 2023, *58* (1), 23–42.

Quansah, Tommy Kweku, Markus Lang, and Bernd Frick, "Racial Bias in the Stands? Investigating Customer-Based Discrimination in European Soccer," *Applied Economics*, 2024, *56* (54), 7050–7064.

Sky Sports, "Police Investigating Racist Abuse of England Players after Euro 2020 Final Make 11 Arrests," https://www.skysports.com/football/news/12016/12372880/polic e-investigating-racist-abuse-of-england-players-after-euro-2020-final-make-11-arrests 2021. Accessed February 23, 2026.

_ , "Jess Carter: England Defender Speaks after Stepping Back from Social Media due to Racism during Euro 2025," https://www.skysports.com/football/news/12040/13400457/jess-carter-england-defender-speaks-after-stepping-back-from-social-media-due-to-racism-during-euro-2025 2025. Accessed February 23, 2026.

Storm, Rasmus K., Cornel Nesseler, Marthe Holum, Andreas Nygaard, and Tor Georg Jakobsen, "Ethnic Discrimination in Scandinavia: Evidence from a Field Experiment in Women's Amateur Soccer," *Humanities and Social Sciences Communications*, 2023, *10* (1), 230.

Szymanski, Stefan, "A Market Test for Discrimination in the English Professional Soccer Leagues," *Journal of Political Economy*, 2000, *108* (3), 590–603.

The Guardian, "Former Spanish PM Accused of Racism in Remarks about French Football Team," https://www.theguardian.com/world/2026/jul/12/spain-mariano-rajoy-accused-of-racism-french-football-team-world-cup 2026. Accessed July 12, 2026.

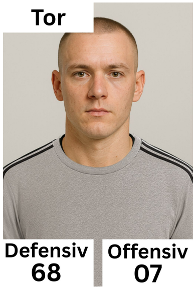
Wozniak, David and Timothy MacNeill, "Racial Discrimination in the Lab: Evidence of Statistical and Taste-Based Discrimination," *Journal of Behavioral and Experimental Economics*, 2020, *85*, 101512.

Zhang, Letian, "A Fair Game? Racial Bias and Repeated Interaction Between NBA Coaches and Players," *Administrative Science Quarterly*, 2017, *62* (4), 603–625.

A. Appendix: Details on the Experiment

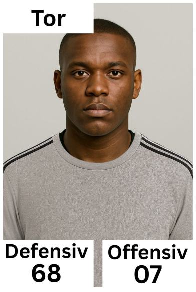
A.1. Players

A.1.1. Goalkeepers



(a) Generic team White.

(b) National team White.



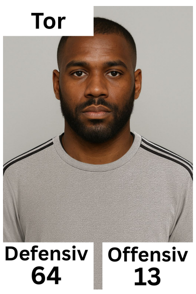
(c) Generic team Black.

(d) National team Black.

Figure 11: Best goalkeeper.

(a) Generic team White.

(b) National team White.

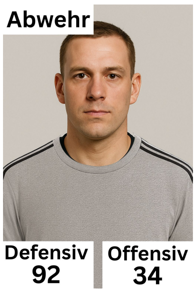


(c) Generic team Black.

(d) National team Black.

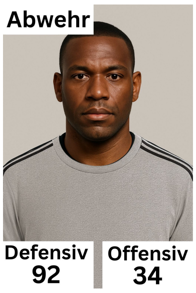
Figure 12: Second-best goalkeeper.

A.1.2. Defenders



(a) Generic team White.

(b) National team White.



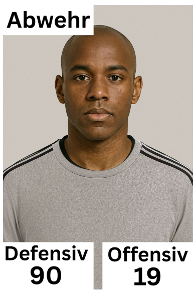
(c) Generic team Black.

(d) National team Black.

Figure 13: Best defender.

(a) Generic team White.

(b) National team White.



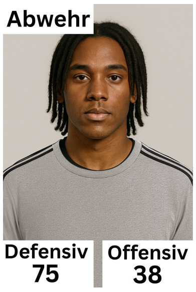
(c) Generic team Black.

(d) National team Black.

Figure 14: Second-best defender.

(a) Generic team White.

(b) National team White.



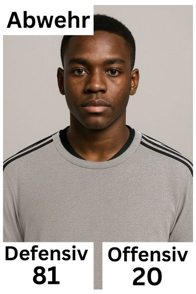
(c) Generic team Black.

(d) National team Black.

Figure 15: Third-best defender.

(a) Generic team White.

(b) National team White.



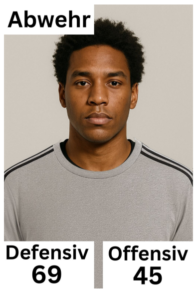
(c) Generic team Black.

(d) National team Black.

Figure 16: Third-worst defender.

(a) Generic team White.

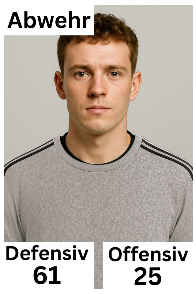
(b) National team White.



(c) Generic team Black.

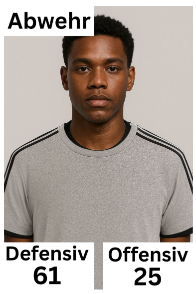
(d) National team Black.

Figure 17: Second-worst defender.



(a) Generic team White.

(b) National team White.



(c) Generic team Black.

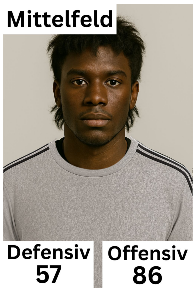
(d) National team Black.

Figure 18: Worst defender.

A.1.3. Midfielders

(a) Generic team White.

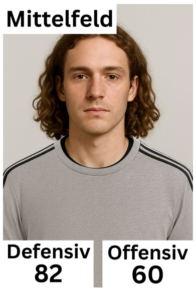
(b) National team White.



(c) Generic team Black.

(d) National team Black.

Figure 19: Best midfielder.



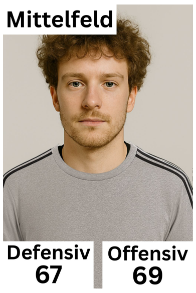
(a) Generic team White.

(b) National team White.

(c) Generic team Black.

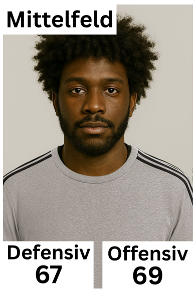
(d) National team Black.

Figure 20: Second-best midfielder.



(a) Generic team White.

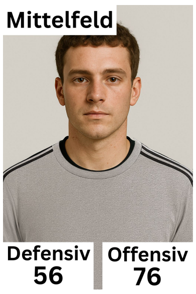
(b) National team White.



(c) Generic team Black.

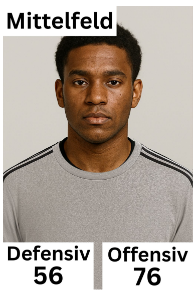
(d) National team Black.

Figure 21: Third-best midfielder.



(a) Generic team White.

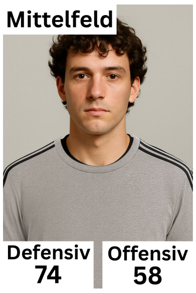
(b) National team White.



(c) Generic team Black.

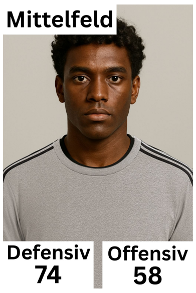
(d) National team Black.

Figure 22: Fourth-best midfielder.



(a) Generic team White.

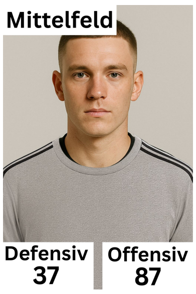
(b) National team White.



(c) Generic team Black.

(d) National team Black.

Figure 23: Fourth-worst midfielder.



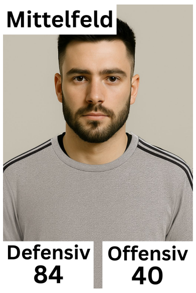
(a) Generic team White.

(b) National team White.

(c) Generic team Black.

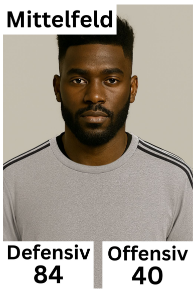
(d) National team Black.

Figure 24: Third-worst midfielder.



(a) Generic team White.

(b) National team White.



(c) Generic team Black.

(d) National team Black.

Figure 25: Second-worst midfielder.

(a) Generic team White.

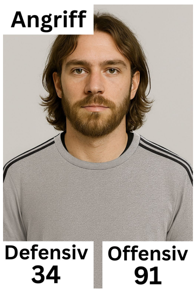
(b) National team White.

(c) Generic team Black.

(d) National team Black.

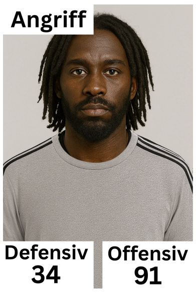
Figure 26: Worst midfielder.

A.1.4. Attackers



(a) Generic team White.

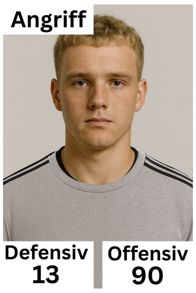
(b) National team White.



(c) Generic team Black.

(d) National team Black.

Figure 27: Best attacker.



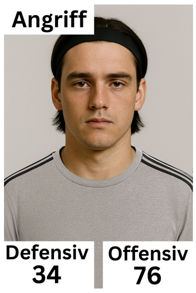
(a) Generic team White.

(b) National team White.

(c) Generic team Black.

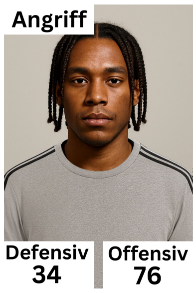
(d) National team Black.

Figure 28: Second-best attacker.



(a) Generic team White.

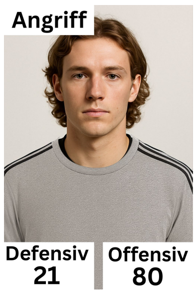
(b) National team White.



(c) Generic team Black.

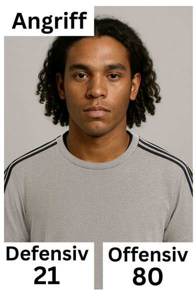
(d) National team Black.

Figure 29: Third-best attacker.



(a) Generic team White.

(b) National team White.



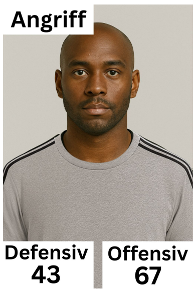
(c) Generic team Black.

(d) National team Black.

Figure 30: Third-worst attacker.

(a) Generic team White.

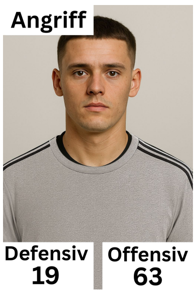
(b) National team White.



(c) Generic team Black.

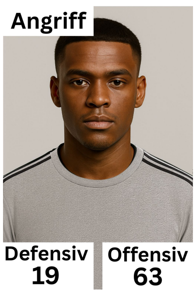
(d) National team Black.

Figure 31: Second-worst attacker.



(a) Generic team White.

(b) National team White.



(c) Generic team Black.

(d) National team Black.

Figure 32: Worst attacker.

A.2. Instructions

Welcome to the study!

This study takes approximately 15 minutes and is funded by the University of Bayreuth. Your participation is voluntary, and you may withdraw at any time without providing a reason. Ending the study prematurely will result in exclusion from all payments but will otherwise have no consequences for you. For your participation, you will receive a **fixed payment of £2.50 (approximately 3.00 €)**. In addition, you can earn a **substantial additional payment**, the amount of which depends on your own decisions. It is therefore important that you **read the following instructions carefully**. Your total earnings will be credited to your Prolific account upon completion of the study. Please participate in a neutral and quiet environment in which you will not be distracted. Please also switch off all mobile devices of any kind (e.g., mobile phone, smartwatch) and put them, as well as any other items that are not required, aside. Please use only the computer programs and functions intended for the study. As part of this study, your responses and decisions, as well as your Prolific ID, will be processed for the purpose of assigning any potential additional payment. The study is hosted using the service provider Heroku (Salesforce, Inc., USA). This may involve the transfer of data to the USA. The data collected will be used exclusively for scientific research purposes. The privacy policy provided at the bottom of this page applies. Detailed instructions for this study will shortly be displayed on your screen. Please read them carefully. You will then be asked comprehension questions about these instructions. If you answer **the same comprehension question incorrectly twice**, you will still receive the fixed payment of £2.50 credited to your Prolific account upon completion of the study, but you will automatically be excluded from all potential additional payments.

By participating, I agree that my responses and decisions, as well as my Prolific ID, may be processed. I also consent to the transfer of data to Heroku (Salesforce, Inc., USA). I understand that I may withdraw my consent at any time—including after submitting my responses—without providing a reason and with effect for the future. In the event of withdrawal, all personal data will be deleted or blocked; research data that have already been anonymized can no longer be linked to my profile.

☐ I have carefully read the introductory text above, taken note of the rules of the study and the privacy policy provided below, and expressly agree to participate and to the associated processing of my data.

Click "Next" to proceed to the instructions.

Next

Instructions: Soccer Lineups

National-Team Treatment

Imagine that, for the upcoming international matches, you take on the role of head coach of the

German men's national soccer team.

Over the past few days, you and your coaching staff have reviewed the squad and discussed possible tactics. It is now time to make the final decision: You must select the starting lineup for the matches. Note: The players shown are **entirely fictional** and are used solely for the purposes of this experiment. Soccer expertise is not required to participate successfully.

Generic-Team Treatment

In the following, your task is to select the starting lineup of a fictional soccer team. You have a squad of 22 players available. Note: The players shown are **entirely fictional** and are used solely for the purposes of this experiment. Soccer expertise is not required to participate successfully.

Joint Instructions

Example of a fictional soccer player:

Each player is identified by their **position**, which is displayed in the upper-left corner. The possible positions are **goalkeeper, defense, midfield, and attack**. Each player also has two attribute values: **Defense** in the lower-left corner and **Offense** in the lower-right corner. Based on these values, each player is assigned an **overall rating**. The attribute values receive different weights depending on the player's position:

1. **Goalkeeper:** The overall rating is determined exclusively by the defensive value (**100%**). The offensive value is not taken into account (**0%**).

2. **Defense:** The defensive value receives greater weight. The overall rating consists of **72%** of the defensive value and **28%** of the offensive value.

3. **Midfield:** Both values are almost equally important. The overall rating consists of **48%** of the defensive value and **52%** of the offensive value.

4. **Attack:** The offensive value receives greater weight. The overall rating consists of **28%** of the defensive value and **72%** of the offensive value.

Decision

Your **task in this study** is to select a starting lineup of **11 players** from a squad of **22 players**. To create the starting lineup, you will be guided through the different positions, goalkeeper, defense, midfield, and attack, one after another, and will select the players whom you would field in each position. Different numbers of players are available in the squad for the different positions:

1. **Goalkeeper:** You select **one out of two players**.

2. **Defense:** You select **three out of six players**.

3. **Midfield:** You select **four out of eight players**.

4. **Attack:** You select **three out of six players**.

The final starting lineup therefore corresponds to a 3-4-3 formation. At the end, your selected starting lineup will be displayed in full, and **you may make adjustments if necessary**.

Payment

Once you have completed your player selection, your starting lineup will receive a **team rating**. To calculate this rating, the **overall ratings of all 11 selected players are added together and divided by 11**. In other words, the average rating determines the team rating. Your team will then compete in **10 matches** against fictional comparison teams. These comparison teams were **created in advance from the same squad** and are identical for all participants. The following applies to each match: If your team has a **higher team rating** than the comparison team, **you win**. For **each victory**, you will receive an **additional £0.50**. You can therefore earn up to an additional £**5.00** in total.

Comprehension Questions

Please answer the following comprehension questions.

1. What is your task in this study?

◯ To select the lineup that I consider the worst.

◯ To select the lineup that I consider the best.

◯ To select a random lineup.

2. Which statement is correct?

◯ For a defensive player, the offensive value receives greater weight.

◯ For a defensive player, the defensive value receives greater weight.

Click "Next" to check your answers.

Next

B. Empirical Strategy

This section describes how we test our hypotheses using the experimental data. The unit of observation is a participant–player profile. Each participant evaluates the same set of $J = 22$ fictional soccer players and selects a starting lineup of 11 players. Our key manipulation is a randomized visual cue of race: each player is shown either with a portrait signaling a Black appearance or a White appearance. Let i index participants and j index players.

B.1. Testing Hypothesis 1: Selection Differences by Race

Hypothesis 1 asks whether, holding player quality fixed, the Black version of a player is selected at a different rate than the White version. Let $selected_{ij} \in \{0, 1\}$ be an indicator equal to one if participant i selects player j into the starting lineup, and zero otherwise. Let $black_{ij} \in \{0, 1\}$ be an indicator equal to one if player j is displayed as Black to participant i, and zero otherwise. To obtain player-specific selection rates and player-specific Black–White differences, we estimate

$$selected_{ij} = \alpha_j + \delta_j \cdot black_{ij} + u_{ij}, \tag{1}$$

where α_j captures the selection probability of player j when displayed as White (the baseline version), and δ_j is the corresponding Black–White difference for player j. In practice, this specification is implemented as a no-intercept regression with player fixed effects and player-specific interactions with $black_{ij}$. Standard errors are clustered at the participant level to account for within-participant correlation across the J player evaluations.

We present the results for Hypothesis 1 as player-specific selection rates for White and Black displays and report the corresponding Black–White differences. We then aggregate the analysis by player category as described next.

A key design feature is that some choices are intended to be relatively easy (very strong or very weak players), while other choices are deliberately close to a knife-edge ("players at the margin"). To formalize this, we classify each player into one of three mutually exclusive categories:

$$\tau_j \in \{\texttt{in}, \texttt{marginal}, \texttt{out}\},$$

where `in` denotes the best players, `out` the worst players, and `marginal` the near-substitute players at the selection threshold. Pooling by category serves two purposes: (i) it provides a compact test of whether race-related differences are concentrated in marginal decisions, and (ii) it facilitates treatment-effect comparisons by reducing dimensionality.

For category-level analysis, we consider an outcome capturing whether a participant makes the *correct* choice given the objective ranking implied by the performance index.

Concretely, a choice is coded as *correct* if (i) `in` players (clear top performers) are selected, (ii) `out` players (clear low performers) are not selected, and (iii) for `marginal` players (near-substitutes at the selection cutoff), the objectively better marginal players are selected, whereas the objectively worse marginal players are not selected. Let $correct_{ij} \in \{0,1\}$ denote this indicator. We estimate

$$correct_{ij} = \beta_{\tau_j} + \gamma_{\tau_j} \cdot black_{ij} + \varepsilon_{ij}. \tag{2}$$

Here, β_τ is the baseline correctness rate for category τ when the player is displayed as White, and γ_τ is the Black–White difference in correctness for category τ. As above, we cluster standard errors at the participant level. For robustness, we re-estimate the model while controlling for the full set of participant characteristics measured in the ex-post questionnaire.

B.2. Testing Hypothesis 2: Racial Cues in Marginal Choices

Hypothesis 2 focuses on the subset of decisions in which the performance index is deliberately very similar across two players. In these `marginal` comparisons, participants face a binary decision between an objectively better player and an objectively worse player, but the quality gap is small enough that the correct choice is difficult to infer from mental arithmetic. This setting is designed to increase the scope for non-performance information—in particular, the randomized racial cue—to affect decisions.

Let i index participants and m index marginal-player pairs. Define $correct_{im} \in \{0,1\}$ as an indicator equal to one if participant i selects the objectively better player from pair m, and zero otherwise. For each marginal-player pair m, we observe the racial appearances of the objectively better and objectively worse players. For each participant, we classify each pair into one of four race configurations:

$$c_{im} \in \{\texttt{Black/Black},\ \texttt{Black/White},\ \texttt{White/Black},\ \texttt{White/White}\},$$

where the first entry refers to the racial appearance of the *better* player and the second entry to that of the *worse* player. For example, `White/Black` denotes a marginal-player pair in which the better player is displayed as White and the worse player as Black.

To compare performance across configurations, we estimate

$$correct_{im} = \theta_{c_{im}} + \varepsilon_{im}. \tag{3}$$

Each coefficient θ_c can be interpreted as the correctness rate (i.e., the probability of choosing the better player) for configuration c. We cluster standard errors at the participant level to account for multiple marginal choices contributed by the same participant.[12]

[12] We restrict the sample to comparisons in which exactly one of the two players at the margin is selected,

Hypothesis 2 implies that, if subtle racial bias affects decisions on the margin, accuracy should be highest when the better player is displayed as White and lowest when the better player is displayed as Black. In terms of (3), this corresponds to the orderings

$$\theta_{\texttt{WB}} > \theta_{\texttt{BB}} > \theta_{\texttt{BW}} \quad \text{and} \quad \theta_{\texttt{WB}} > \theta_{\texttt{WW}} > \theta_{\texttt{BW}},$$

with the strongest contrast given by $\theta_{\texttt{WB}} - \theta_{\texttt{BW}}$. For robustness, we re-estimate the model while controlling for the full set of participant characteristics measured in the ex-post questionnaire.

B.3. Testing Hypothesis 3: Effects of the National Team Framing

To study whether the national-team framing changes behavior, we assign participants to either the generic-team or the national-team treatment. Let w_i be an indicator equal to one if participant i is assigned to the national-team treatment and zero otherwise. We allow both baseline correctness rates and Black–White differences to vary by treatment status by estimating

$$correct_{ij} = \beta_{\tau_j} + \gamma_{\tau_j} \cdot black_{ij} + \lambda_{\tau_j} \cdot w_i + \kappa_{\tau_j} \cdot black_{ij} \cdot w_i + \eta_{ij}. \tag{4}$$

For each player category τ, γ_τ is the Black–White difference in the generic-team condition, while $\gamma_\tau + \kappa_\tau$ is the corresponding difference in the national-team condition. Thus, κ_τ captures how the national-team framing changes the Black–White difference in correctness. We again cluster standard errors at the participant level.

To test whether the national-team framing amplifies or attenuates the configuration effects for players at the margin, we allow configuration-specific correctness rates to differ across treatment conditions. Let again w_i indicate assignment to the national-team treatment. We estimate

$$correct_{im} = \theta_{c_{im}} + \phi_{c_{im}} \cdot w_i + \nu_{im}. \tag{5}$$

For each racial configuration c, θ_c denotes the correctness rate in the generic-team condition, while ϕ_c captures the corresponding effect of the national-team framing. Thus, the correctness rate in the national-team condition is $\theta_c + \phi_c$.

We report and compare the implied correctness rates across configurations separately for the generic-team and national-team conditions. To assess whether the national-team framing affects behavior, we examine the estimated treatment differences, ϕ_c, for each configuration.

dropping cases where both or neither player is chosen. As a robustness check, we re-estimate the specification on the full set of marginal observations using $selected_{im}$ as the dependent variable, which indicates whether the better player at the margin is selected.

For robustness, we re-estimate both models while controlling for the full set of participant characteristics measured in the ex-post questionnaire.

C. Appendix: Data and Covariate Balance Test

C.1. Data

Table 3: Overview of metric variables included in the ex-post questionnaire.

Variable	**Min.**	**Median**	**Mean**	**Std. Dev.**	**Max.**
Age	18.00	29.00	30.85	9.13	74.00
National team interest	1.00	4.00	4.66	2.93	10.00
General soccer interest	1.00	4.00	4.84	3.07	10.00

Table 4: Overview of categorical variables included in the ex-post questionnaire.

Variable	**Categories**	**# observations**
Gender	Female	421
	Male	768
	non-binary	11
		Total = 1200
Origin	Germany	1082
	Austria or Switzerland	3
	Non-German-speaking Europe	59
	None of the above	56
		Total = 1200
Party affiliation	LINKE	319
	GRUENE	284
	SPD	176
	BSW	58
	FDP	120
	CDU/CSU	143

Continued on next page

Variable	Categories	# observations
	AfD	100
		Total = 1200
National team statement	Fully disagree	516
	Disagree	210
	Neutral	418
	Agree	43
	Fully agree	13
		Total = 1200
Coaching experience	None	1078
	Yes, but other sports	84
	Yes, in soccer	38
		Total = 1200
Actively playing soccer	No	673
	Only during leisure time	267
	Yes, in a soccer club	260
		Total = 1200
Playing computer games	Regularly	200
	Occasionally	497
	No	503
		Total = 1200
Migrational background	Migrant	185
	Parents are migrants	206
	Grandparents are migrants	33
	None	776
		Total = 1200
Computational skills	Very high	196
	High	535
	Average	391
	Low	69

Continued on next page

Variable	Categories	# observations
	Very low	9
		Total = 1200

C.2. Covariate Balance Test

Table 5: Covariate Balance Across the Generic- and National-Team Conditions

Variable	Generic	National	Diff. (N–G)	p-value
Party affiliation				
LINKE	28.3	24.8	-3.5	0.170
GRUENE	23.0	24.3	1.3	0.587
SPD	15.3	14.0	-1.3	0.514
BSW	5.2	4.5	-0.7	0.591
FDP	8.5	11.5	3.0	0.083
CDU/CSU	11.7	12.2	0.5	0.789
AfD	8.0	8.7	0.7	0.676
Age	30.30	31.40	1.10	0.037
Gender				
Female	39.0	31.2	-7.8	0.004
Male	60.2	67.8	7.7	0.006
Non-binary	0.8	1.0	0.2	0.762
National team statement				
Fully disagree	45.7	40.3	-5.3	0.062
Disagree	16.8	18.2	1.3	0.544
Neutral	32.8	36.8	4.0	0.146
Agree	4.2	3.0	-1.2	0.277
Fully agree	0.5	1.7	1.2	0.051
National team interest	4.53	4.79	0.26	0.127
Coaching experience				
None	89.8	89.8	0.0	1.000
Yes, but other sports	6.7	7.3	0.7	0.651
Yes, in soccer	3.5	2.8	-0.7	0.510
Actively playing soccer				
No	56.2	56.0	-0.2	0.954
Only during leisure time	22.2	22.3	0.2	0.945
Yes, in a soccer club	21.7	21.7	0.0	1.000

Continued on next page

Variable	Generic	National	Diff. (N–G)	p-value
Soccer interest	4.64	5.04	0.40	0.026
Playing computer games				
Regularly	14.0	19.3	5.3	0.013
Occasionally	42.8	40.0	-2.8	0.320
No	43.2	40.7	-2.5	0.381
Origin				
Germany	89.2	91.2	2.0	0.245
Austria or Switzerland	0.3	0.2	-0.2	0.564
Non-German-speaking Europe	6.3	3.5	-2.8	0.023
None of the listed regions	4.2	5.2	1.0	0.412
Migrational background				
Migrant	15.7	15.2	-0.5	0.811
Parents are migrants	20.0	14.3	-5.7	0.009
Grandparents are migrants	2.8	2.7	-0.2	0.860
None	61.5	67.8	6.3	0.022
Computational skills				
Very high	15.7	17.0	1.3	0.533
High	44.5	44.7	0.2	0.954
Average	33.0	32.2	-0.8	0.758
Low	6.0	5.5	-0.5	0.710
Very low	0.8	0.7	-0.2	0.738
Joint orthogonality test (HC1 robust Wald): F=1.64, p=0.032				

Notes: Generic and National columns report means (or percentages for categorical levels). Diff. is National minus Generic. Individual p-values come from HC1-robust regressions of each covariate indicator (or numeric covariate) on the treatment indicator. The joint test reports an HC1-robust Wald F-test that all covariate coefficients are jointly zero. Covariates that may be affected by the treatment are excluded, namely national-team statement, national team interest, and soccer interest.

D. Appendix: Experimental Results and Additional Regressions

Table 6: Player-Specific Selection Rates and Black–White Differences

	(1)
Goal: Player 1	0.868***
	(0.014)
Goal: Player 1 × Black	0.051***
	(0.018)
Goal: Player 2	0.118***
	(0.013)
Goal: Player 2 × Black	-0.024
	(0.018)
Defense: Player 1	0.966***
	(0.007)
Defense: Player 1 × Black	0.002
	(0.010)
Defense: Player 2	0.889***
	(0.013)
Defense: Player 2 × Black	-0.002
	(0.018)
Defense: Player 3	0.452***
	(0.020)
Defense: Player 3 × Black	0.033
	(0.029)
Defense: Player 4	0.541***
	(0.020)
Defense: Player 4 × Black	0.004
	(0.029)
Defense: Player 5	0.093***
	(0.012)
Defense: Player 5 × Black	0.046**
	(0.018)
Defense: Player 6	0.013***
	(0.005)
Defense: Player 6 × Black	0.008

Continued on next page

	(1)
	(0.007)
Midfield: Player 1	0.824***
	(0.015)
Midfield: Player 1 × Black	0.032
	(0.021)
Midfield: Player 2	0.827***
	(0.015)
Midfield: Player 2 × Black	-0.019
	(0.022)
Midfield: Player 3	0.810***
	(0.016)
Midfield: Player 3 × Black	0.023
	(0.022)
Midfield: Player 4	0.607***
	(0.020)
Midfield: Player 4 × Black	-0.010
	(0.028)
Midfield: Player 5	0.401***
	(0.020)
Midfield: Player 5 × Black	0.015
	(0.028)
Midfield: Player 6	0.202***
	(0.016)
Midfield: Player 6 × Black	0.002
	(0.023)
Midfield: Player 7	0.159***
	(0.015)
Midfield: Player 7 × Black	0.002
	(0.021)
Midfield: Player 8	0.154***
	(0.015)
Midfield: Player 8 × Black	-0.012
	(0.021)
Attack: Player 1	0.978***
	(0.006)

Continued on next page

	(1)
Attack: Player 1 × Black	0.001
	(0.008)
Attack: Player 2	0.908***
	(0.012)
Attack: Player 2 × Black	-0.018
	(0.017)
Attack: Player 3	0.445***
	(0.020)
Attack: Player 3 × Black	0.002
	(0.029)
Attack: Player 4	0.600***
	(0.020)
Attack: Player 4 × Black	-0.029
	(0.028)
Attack: Player 5	0.066***
	(0.010)
Attack: Player 5 × Black	0.019
	(0.015)
Attack: Player 6	0.008**
	(0.004)
Attack: Player 6 × Black	0.015**
	(0.007)
Observations	26,400
R^2	0.732

Notes: The dependent variable is an indicator equal to one if the player is selected and zero otherwise. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 7: Decision Accuracy by Player Category and Racial Appearence

	(1)	(2)
In	0.883***	0.790***
	(0.005)	(0.039)
Marginal	0.493***	0.400***

Continued on next page

	(1)	(2)
	(0.011)	(0.040)
Out	0.899***	0.806***
	(0.005)	(0.039)
In × Black	0.010	0.012*
	(0.006)	(0.006)
Marginal × Black	0.006	0.006
	(0.012)	(0.012)
Out × Black	-0.008	-0.007
	(0.006)	(0.006)
Games experience: Occasionally		-0.001
		(0.012)
Games experience: No		-0.012
		(0.015)
National-team statement: Disagree		-0.011
		(0.010)
National-team statement: Neutral		-0.023***
		(0.008)
National-team statement: Agree		-0.023
		(0.021)
National-team statement: Fully agree		0.010
		(0.040)
Party: BSW		0.034*
		(0.020)
Party: CDU/CSU		0.009
		(0.016)
Party: FDP		0.019
		(0.017)
Party: GRUENE		0.054***
		(0.015)
Party: LINKE		0.022
		(0.016)
Party: SPD		0.013
		(0.016)
Gender: Male		0.043***

Continued on next page

	(1)	(2)
		(0.009)
Gender: Non-binary		-0.024
		(0.033)
Age		-0.001**
		(0.000)
Origin: Austria or Switzerland		0.131**
		(0.028)
Origin: Non german-speaking Europe		0.100*
		(0.031)
Origin: Other		0.110*
		(0.033)
Computuational skills: High		-0.026**
		(0.011)
Computational skills: Average		-0.063***
		(0.012)
Computational skills: Low		-0.083***
		(0.016)
Computational skills: Very low		-0.087*
		(0.043)
General soccer interest		-0.002
		(0.002)
National team interest		-0.002
		(0.002)
Soccer active		0.014**
		(0.006)
Coaching experience		-0.003
		(0.009)
Observations	26,400	26,400
R^2	0.823	0.826

Notes: The dependent variable is whether the player classification was correct. Model (1) includes no participant-level controls. Model (2) includes all participant-level controls. The variables soccer active and coaching experience are binary variables summarizing the categories into yes and no. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 8: Correctness Rates by Race Configuration

	(1)	(2)
Black–Black	0.511***	0.218
	(0.019)	(0.149)
Black–White	0.485***	0.200
	(0.019)	(0.147)
White–Black	0.500***	0.209
	(0.019)	(0.149)
White–White	0.489***	0.200
	(0.019)	(0.148)
Games: Occasionally		0.012
		(0.036)
Games: No		0.028
		(0.045)
National-team statement: Disagree		-0.007
		(0.031)
National-team statement: Neutral		0.004
		(0.026)
National-team statement: Agree		-0.022
		(0.059)
National-team statement: Fully agree		0.086
		(0.103)
Party: BSW		-0.014
		(0.060)
Party: CDU/CSU		-0.016
		(0.047)
Party: FDP		0.067
		(0.050)
Party: GRUENE		0.108**
		(0.044)
Party: LINKE		0.026
		(0.044)
Party: SPD		0.013
		(0.046)
Gender: Male		0.046*

Continued on next page

	(1)	(2)
		(0.026)
Gender: Non-binary		0.005
		(0.101)
Age		-0.001
		(0.001)
Origin: Austria and Switzerland		0.312
		(0.127)
Origin: Non-German-speaking Europe		0.263
		(0.136)
Origin: Other		0.278
		(0.137)
Computational skills: High		-0.066**
		(0.032)
Computational skills: Average		-0.146***
		(0.035)
Computational skills: Low		-0.209***
		(0.048)
Computational skills: Very low		0.042
		(0.134)
General soccer interest		0.001
		(0.008)
National-team interest		-0.001
		(0.007)
Soccer active		0.031*
		(0.017)
Coaching experience		-0.010
		(0.027)
Observations	3,100	3,100
R^2	0.496	0.512

Notes: The dependent variable indicates whether the better marginal player was selected. Only the cases in which exactly one of the two players was selected are included. Model (1) includes configuration indicators only. Model (2) additionally includes all participant-level controls. The variables soccer active and coaching experience are binary variables summarizing the categories into yes and no. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 9: Player Selection by Race Configuration

	(1)	(2)
Black–Black	0.527***	0.566***
	(0.017)	(0.024)
Black–White	0.492***	0.495***
	(0.018)	(0.024)
White–Black	0.512***	0.498***
	(0.018)	(0.026)
White–White	0.490***	0.489***
	(0.017)	(0.024)
Black–Black × National-team		-0.074**
		(0.034)
Black–White × National-team		-0.005
		(0.036)
White–Black × National-team		0.028
		(0.036)
White–White × National-team		0.002
		(0.035)
Observations	3,600	3,600
R^2	0.506	0.507

Notes: The dependent variable indicates whether the better marginal player was selected. All observations are included. Model (1) includes configuration indicators only. Model (2) adds national-team interactions. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 10: Decision Accuracy by Player Category, Racial Appearence, and Treatment

	(1)	(2)
In	0.883***	0.789***
	(0.007)	(0.039)
Marginal	0.488***	0.396***
	(0.016)	(0.040)
Out	0.900***	0.806***
	(0.007)	(0.039)
In × Black	0.008	0.010

Continued on next page

	(1)	(2)
	(0.009)	(0.009)
Marginal × Black	0.024	0.022
	(0.017)	(0.017)
Out × Black	-0.008	-0.005
	(0.008)	(0.008)
In × National-team	-0.000	-0.003
	(0.010)	(0.010)
Marginal × National-team	0.010	0.005
	(0.022)	(0.022)
Out × National-team	-0.001	-0.003
	(0.009)	(0.009)
In × Black × National-team	0.005	0.003
	(0.013)	(0.013)
Marginal × Black × National-team	-0.035	-0.031
	(0.024)	(0.024)
Out × Black × National-team	0.000	-0.003
	(0.012)	(0.012)
Games: Occasionally		-0.001
		(0.012)
Games: No		-0.012
		(0.015)
National-team statement: Disagree		-0.011
		(0.010)
National-team statement: Neutral		-0.023***
		(0.008)
National-team statement: Agree		-0.023
		(0.021)
National-team statement: Fully agree		0.010
		(0.039)
Party: BSW		0.034*
		(0.020)
Party: CDU/CSU		0.009
		(0.016)
Party: FDP		0.020
		(0.017)

Continued on next page

	(1)	(2)
Party: GRUENE		0.054***
		(0.016)
Party: LINKE		0.022
		(0.016)
Party: SPD		0.013
		(0.016)
Gender: Male		0.043***
		(0.009)
Gender: Non-binary		-0.024
		(0.034)
Age		-0.001**
		(0.000)
Origin: Austria and Switzerland		0.132**
		(0.027)
Origin: Non-German-speaking Europe		0.100*
		(0.030)
Origin: Other		0.110*
		(0.032)
Computational skills: High		-0.026**
		(0.011)
Computational skills: Average		-0.063***
		(0.012)
Computational skills: Low		-0.083***
		(0.016)
Computational skills: Very low		-0.087*
		(0.043)
General soccer interest		-0.002
		(0.002)
National-team interest		-0.002
		(0.002)
Soccer active		0.013**
		(0.006)
Coaching experience		-0.003
		(0.009)

Continued on next page

	(1)	(2)
Observations	26,400	26,400
R^2	0.824	0.826

Notes: The dependent variable indicates whether the player classification was correct. Model (1) includes no participant-level controls. Model (2) includes all participant-level controls. The variables soccer active and coaching experience are binary variables summarizing the categories into yes and no. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 11: Correctness Rates by Race Configuration and Treatment

	(1)	(2)
Black–Black	0.558***	0.258
	(0.027)	(0.150)
Black–White	0.493***	0.202
	(0.026)	(0.148)
White–Black	0.473***	0.178
	(0.028)	(0.150)
White–White	0.479***	0.193
	(0.026)	(0.149)
Black–Black × National-team	-0.091**	-0.090**
	(0.038)	(0.038)
Black–White × National-team	-0.018	-0.020
	(0.038)	(0.037)
White–Black × National-team	0.051	0.044
	(0.039)	(0.039)
White–White × National-team	0.020	-0.002
	(0.037)	(0.037)
Games: Occasionally		0.011
		(0.036)
Games: No		0.028
		(0.045)
National-team statement: Disagree		-0.006
		(0.031)
National-team statement: Neutral		0.004
		(0.026)

Continued on next page

	(1)	(2)
National-team statement: Agree		-0.025
		(0.059)
National-team statement: Fully agree		0.087
		(0.103)
Party: BSW		-0.013
		(0.060)
Party: CDU/CSU		-0.014
		(0.046)
Party: FDP		0.067
		(0.050)
Party: GRUENE		0.109**
		(0.044)
Party: LINKE		0.029
		(0.044)
Party: SPD		0.014
		(0.046)
Gender: Male		0.048*
		(0.027)
Gender: Non-binary		0.012
		(0.103)
Age		-0.001
		(0.001)
Origin: Austria and Switzerland		0.316
		(0.128)
Origin: Non-German-speaking Europe		0.261
		(0.136)
Origin: Other		0.281
		(0.138)
Computational skills: High		-0.064**
		(0.032)
Computational skills: Average		-0.145***
		(0.035)
Computational skills: Low		-0.210***
		(0.048)
Computational skills: Very low		0.041

Continued on next page

	(1)	(2)
		(0.137)
General soccer interest		0.001
		(0.008)
National-team interest		-0.001
		(0.007)
Soccer active		0.029*
		(0.017)
Coaching experience		-0.010
		(0.027)
Observations	3,100	3,100
R^2	0.497	0.514

Notes: The dependent variable indicates whether the better marginal player was selected. Model (1) includes no participant-level controls. Model (2) includes all participant-level controls. The variables soccer active and coaching experience are binary variables summarizing the categories into yes and no. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 12: Determinants of Performance (number of wins)

	(1)
Party: GRUENE	0.464
	(0.311)
Party: SPD	-0.218
	(0.358)
Party: BSW	0.124
	(0.542)
Party: FDP	-0.417
	(0.428)
Party: CDU/CSU	-0.591
	(0.393)
Party: AfD	-0.875*
	(0.453)
Age	-0.021*
	(0.012)

Continued on next page

	(1)
Gender: Male	1.323***
	(0.270)
Gender: Non-binary	-0.390
	(1.113)
National-team statement: Disagree	-0.234
	(0.306)
National-team statement: Neutral	-0.774***
	(0.259)
National-team statement: Agree	-1.003
	(0.626)
National-team statement: Fully agree	0.304
	(1.164)
Soccer active: Leisure time	0.187
	(0.292)
Soccer active: Yes	0.360
	(0.348)
Coaching experience: Other sports	-0.127
	(0.409)
Coaching experience: Soccer	-0.142
	(0.679)
General soccer interest: High	-0.616*
	(0.315)
General soccer interest: Medium	-0.738**
	(0.303)
Games experience: Occasionally	-0.238
	(0.369)
Games experience: No	-0.686
	(0.443)
Migration: Parents	0.307
	(0.432)
Migration: Grandparents	1.160
	(0.738)
Migration: No	0.673*
	(0.377)
Origin: Austria and Switzerland	-4.403***

Continued on next page

	(1)
	(0.481)
Origin: Europe	-0.755
	(0.534)
Origin: None	0.092
	(0.598)
Computational skills: High	-0.644*
	(0.332)
Computational skills: Average	-1.880***
	(0.357)
Computational skills: Low	-2.906***
	(0.480)
Computational skills: Very low	-3.106***
	(1.031)
Duration	0.047
	(0.042)
Constant	5.384***
	(0.846)
Observations	1,200
R^2	0.155

Notes: OLS regression with heteroskedasticity-robust (HC1) standard errors in parentheses. The dependent variable is *wins*.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 13: Decision Accuracy by Player Category, Racial Appearence, and Attitude Group

	(1)	(2)
In	0.867***	0.771***
	(0.009)	(0.040)
Marginal	0.503***	0.407***
	(0.018)	(0.043)
Out	0.879***	0.781***
	(0.008)	(0.040)
In × Black	0.002	0.003
	(0.011)	(0.011)

Continued on next page

	(1)	(2)
Marginal × Black	-0.021	-0.021
	(0.019)	(0.019)
Out × Black	0.001	0.005
	(0.010)	(0.010)
In × Reject	0.027**	0.023**
	(0.011)	(0.011)
Marginal × Reject	-0.016	-0.021
	(0.023)	(0.023)
Out × Reject	0.033***	0.031***
	(0.010)	(0.010)
In × Black × Reject	0.013	0.014
	(0.013)	(0.013)
Marginal × Black × Reject	0.044*	0.045*
	(0.024)	(0.024)
Out × Black × Reject	-0.014	-0.019
	(0.012)	(0.012)
Games: Occasionally		-0.002
		(0.012)
Games: No		-0.012
		(0.015)
Party: BSW		0.032
		(0.020)
Party: CDU/CSU		0.006
		(0.016)
Party: FDP		0.018
		(0.017)
Party: GRUENE		0.052***
		(0.015)
Party: LINKE		0.020
		(0.015)
Party: SPD		0.010
		(0.016)
Gender: Male		0.043***
		(0.009)
Gender: Non-binary		-0.020

Continued on next page

	(1)	(2)
		(0.033)
Age		-0.001**
		(0.000)
Origin: Austria and Switzerland		0.128**
		(0.028)
Origin: Non-German-speaking Europe		0.098*
		(0.032)
Origin: Other		0.107*
		(0.033)
Computational skills: High		-0.026**
		(0.011)
Computational skills: Average		-0.063***
		(0.012)
Computational skills: Low		-0.083***
		(0.016)
Computational skills: Very low		-0.088*
		(0.043)
General soccer interest		-0.002
		(0.002)
National-team interest		-0.002
		(0.002)
Soccer active		0.013**
		(0.006)
Coaching experience		-0.003
		(0.009)
Observations	26,400	26,400
R^2	0.824	0.826

Notes: The dependent variable indicates whether the player classification was correct. Model (1) includes no participant-level controls. Model (2) includes participant-level controls. The variables soccer active and coaching experience are binary variables summarizing the categories into yes and no. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 14: Correctness Rates by Race Configuration and Attitude Group

	(1)	(2)
Black–Black	0.486***	0.211
	(0.030)	(0.150)
Black–White	0.487***	0.217
	(0.032)	(0.150)
White–Black	0.497***	0.222
	(0.029)	(0.151)
White–White	0.497***	0.228
	(0.029)	(0.150)
Black–Black × Reject	0.042	0.029
	(0.039)	(0.039)
Black–White × Reject	-0.004	-0.013
	(0.040)	(0.041)
White–Black × Reject	0.006	-0.007
	(0.039)	(0.039)
White–White × Reject	-0.013	-0.031
	(0.038)	(0.039)
Games: Occasionally		0.011
		(0.036)
Games: No		0.028
		(0.045)
Party: BSW		-0.019
		(0.059)
Party: CDU/CSU		-0.021
		(0.046)
Party: FDP		0.065
		(0.049)
Party: GRUENE		0.105**
		(0.043)
Party: LINKE		0.023
		(0.043)
Party: SPD		0.008
		(0.045)
Gender: Male		0.046*

Continued on next page

	(1)	(2)
		(0.026)
Gender: Non-binary		0.011
		(0.103)
Age		-0.001
		(0.001)
Origin: Austria and Switzerland		0.308
		(0.127)
Origin: Non-German-speaking Europe		0.258
		(0.135)
Origin: Other		0.272
		(0.137)
Computational skills: High		-0.067**
		(0.032)
Computational skills: Average		-0.146***
		(0.035)
Computational skills: Low		-0.210***
		(0.048)
Computational skills: Very low		0.038
		(0.134)
General soccer interest		0.001
		(0.008)
National-team interest		-0.001
		(0.007)
Soccer active		0.031*
		(0.017)
Coaching experience		-0.011
		(0.027)
Observations	3,100	3,100
R^2	0.496	0.512

Notes: The dependent variable indicates whether the better marginal player was selected. Model (1) includes no participant-level controls. Model (2) includes participant-level controls. The variables soccer active and coaching experience are binary variables summarizing the categories into yes and no. CR2 cluster-robust standard errors at the participant level are reported in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.