Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)

Racial Biases in Social Media Discussions About Soccer Players

Tobias Schumacher[1, 2], **Indira Sen**[1], **Esi Doda**[2], **Claudia Wagner**[2,3,4]

[1]University of Mannheim, Mannheim, Germany
[2]RWTH Aachen University, Aachen, Germany
[3]GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany
[4]Complexity Science Hub, Vienna, Austria
tobias.schumacher@uni-mannheim.de, indira.sen@uni-mannheim.de, esi.doda@rwth-aachen.de, claudia.wagner@gesis.org

Abstract

Recent reports suggest that Black soccer players are often subjected to racial abuse from fans in social media commentary. This study develops a pipeline to collect data from Reddit and Twitter during a four-season period spanning three soccer tournaments, linking posts with mentioned soccer players to investigate the prevalence of racial bias in these posts. Our results highlight subtle forms of racial disparity in social media comments about soccer players. Specifically, we find that Black players tend to receive more toxic comments than White players and that Black players receive more comments related to their physical attributes, in line with previous findings in traditional media coverage. We hope that our research and multiplatform social media datasets help raise awareness of subtle forms of racial disparities in the discussion of athletes and motivate further research to characterize racial bias in soccer.

1 Introduction

Racism still poses a prevalent problem in international sports, particularly in sports as popular as soccer, as exemplified by several recent reports. For example, more than 600 racist incidents have been reported in English soccer to the *Kick It Out* organization in the recent 2024/25 season (Kick It Out 2025). In Germany, a TV broadcaster survey prior to Euro 2024 has found that 21% of Germans would prefer more White players on their national team (Lukiv 2024), while in Spain, the Real Madrid player Vinicius Jr. broke down during a press conference due to the constant racist abuse directed at him (Brewin 2024).

Recent research has also pointed out more implicit forms of racial biases directed towards players of color by commentators, referees, or fans: commentators and fans are more likely to praise Black players for their physicality compared to White players. In turn, the latter are more often credited with 'intelligence' (Campbell and Bebb 2021). Black players are also more likely to receive yellow cards from referees than White players (Magistro and Wack 2023).

Hate speech toward players also occurs on online social media, as recent studies have confirmed, while noting that explicit racist abuse is relatively rare (Vidgen et al. 2022; Cullen and Williams 2023). However, there is also a lack of studies that investigate more implicit racial biases toward soccer players on social media. In this paper, we fill this gap by providing a large-scale analysis of soccer discussions on online social media with respect to racial biases. We specifically focus on discussions about male players competing in the English Premier League during the 2017/18-2020/21 seasons and analyze discussions on Reddit and Twitter. In line with findings from previous research, we investigate the following three research questions.

RQ1: Are non-White players more likely to receive negative comments than White players on online social media?

RQ2: Are non-White players more likely to receive toxic comments than White players on online social media?

RQ3: Are there differences in the frequency of comments related to physicality, technical skills, or mental attributes of players based on their perceived race?

We investigate these questions by collecting millions of comments from Reddit and Twitter, as well as collecting metadata about soccer players from *Transfermarkt*, *WhoScored* and *Soccer Wiki*. We augment these datasets by identifying the perceived race of players via a triangulation approach and identify sentiment and toxicity of comments using state-of-the-art NLP methods. We merge player and comment datasets using a refined matching approach based on named entity recognition and fuzzy string matching, and assign categories of 'soccer supersenses' (physical, technical, mental) to comments by matching an extensive, handmade dictionary, refined further using an LLM-based disambiguation approach. Finally, we apply a combination of regression and matching analysis to investigate the impact of race on the way players are discussed online, while controlling for several confounders such as a player's popularity.

We find modest associations between player race and the toxicity of comments they receive on both platforms, suggesting that subtle racial biases are still present in social media. In line with previous research, we further find that Black players tend to receive more comments relating to their physicality than White players. However, we do not find a significant correlation between player race and comment sentiment. Overall, our results highlight the need for taking a multifaceted and holistic perspective when analyzing online commentary towards soccer players.

In summary, we present a large-scale analysis of social media data regarding implicit racial biases. To facilitate further research, we provide gated access to our dataset and share all our code in a publicly available repository[1].

2 Soccer and Competitions under Study

Soccer is widely considered the world's most popular sport, with an estimated 250 million active players (FIFA 2007), and a worldwide viewership of 571 million people for the recent 2022 World Cup final—making events like the Super Bowl in American Football "pale in comparison" (Richter 2025). This immense global interest has led to the formation of large communities, and naturally, on social media, soccer has also become a huge topic of interest.

In this study, we focus on online discussions related to the English *Premier League*[2], which is the highest-tier English soccer league. It is also the most watched domestic sports league globally, broadcast in 189 countries with an audience exceeding one billion people (Moore 2024). Each season, twenty teams compete in this league, with the three worst-performing teams getting relegated. However, Premier League clubs also compete for other trophies, both nationally and internationally. Of these, we specifically considered the *FA Cup*[3], and the *UEFA Champions League*[4].

The FA Cup is the oldest soccer tournament in the world—a domestic single-elimination tournament played annually in England, organized by the English *Football Association (FA)*, and taking place in concurrence with the regular Premier League season.

The Champions League is the most prestigious club-level competition in the world, organized by *Union of European Football Associations (UEFA)*. Conceptually, the strongest teams in Europe compete in this tournament, typically qualifying by placing in top spots within their domestic league—stronger leagues (as determined by a 5-year-performance ranking) get to send more teams to this tournament. In each of the 2017/18-2020/21 seasons, 32 teams competed in the Champions League, with 4-5 of them from Premier League.

3 Related Work

Next, we summarize existing work related to our study.

Biases in Traditional Media Coverage. There is a long-standing body of research that has analyzed racial biases in traditional media coverage of soccer. More than 20 years ago, McCarthy, Jones, and Potrac (2003) analyzed TV commentary on British soccer matches in the 1997/98 season and found that, while there was no difference in the sentiment toward players of different races, Black players received much more praise related to their physicality compared to White players. This finding has been confirmed in the context of Dutch TV coverage of their national soccer league (van Sterkenburg, Knoppers, and de Leeuw 2012), British TV coverage of the 2018 World Cup (Campbell and Bebb 2021), Spanish (Longas Luque and van Sterkenburg 2022a) and Polish (van Lienden and van Sterkenburg 2023) TV commentary of their national teams, and in British, Canadian, and US-American coverage of the top 4 European leagues (McLoughlin 2023). However, there are also contradictory results: Longas Luque and van Sterkenburg (2022b) found reverse patterns in the way Black players' physicality was discussed on Spanish TV, and van Campenhout, van Lienden, and van Sterkenburg (2023) found that Dutch commentators did not display such biases when talking about players from their own national team. Next to such biases in match commentary, (Principe and van Ours 2022) have also analyzed the existence of biases in ratings of *Serie A* players by Italian newspapers between 2009 and 2018. They found that at the lower end of the ratings distributions, Black players tended to receive poorer ratings than White players.

Biases in Offline Fan Discussions. Several qualitative studies have analyzed offline discourse of soccer fans, finding biases similar to those presented by media commentators. (Müller, van Zoonen, and de Roode 2007) interviewed 20 fans and youth players in Amsterdam and discovered a pattern of "accidental racism", where fans engage in racist behaviors while claiming not to mean it as such. In a focus group study on Dutch media audiences, the stereotype of more athletic black players was reproduced, and while diversity was celebrated, this was in part limited and conditional on the loyalty of multi-ethnic players to the Dutch nation (van Sterkenburg, Peeters, and van Amsterdam 2019). Similarly, the stereotype of the Black athlete has been reproduced in focus group studies with young Polish (van Lienden, van Sterkenburg, and Sommier 2023), Spanish (Longas Luque, Sommier, and van Sterkenburg 2024) fans and English fans, although in the latter, this stereotype has also been challenged. Further, among both the Polish and the English fans, stereotypes about nationalities have been invoked, such as players from Eastern European countries having more fighting spirit or Spanish players having superior technical skills.

Social Media Studies. Despite extensive research on racism in online social media (Bliuc et al. 2018; Matamoros-Fernández and Farkas 2021; Castaño-Pulgarín et al. 2021), relatively few studies focus on sports, or specifically soccer-related online discourse. In an earlier study, Cleland (2014) qualitatively analyzed more than 500 posts on two online message boards related to English lower-tier soccer clubs, finding that online social media offer a new space where racist discourse could move. This finding has been confirmed by Bennet and Jönsson (2017), who identified almost 135,000 discriminatory social media posts related to the Premier League in the 2014/15 season. More recently, Cable, Kilvington, and Mottershead (2022) analyzed 8845 Twitter comments related to an incident in a 2019 Premier League match where Antonio Rüdiger experienced racial abuse from fans, finding largely supportive comments, but also a small minority of abusive or racist comments toward Rüdiger. Similarly, Glynn and Brown (2023) qualitatively analyzed the use of discriminatory humor in the soccer Twitter community, finding that despite widespread offline condemnation, racism remains a problem that has shifted to online social media. In a more quantitative analysis, Miranda et al. (2024) collected 276,231 fan comments related to Portuguese soccer clubs at the beginning of the 2020/2021 season and found a very low incidence of 0.15% of comments

[1]https://github.com/tobiasschumacher/bias-soccer-discussions
[2]https://www.premierleague.com/
[3]https://www.thefa.com/competitions/thefacup/
[4]https://www.uefa.com/uefachampionsleague/

containing hate speech, including racism and xenophobia. Similarly, Vidgen et al. (2022) and Cullen and Williams (2023) conducted large-scale analyses of social media posts on the 2020/21 Premier League season and Euro 2022. Both studies confirm the occurrence of abusive comments toward players but also find that explicit racist abuse against players was relatively rare. Finally, a concurrent work by Alrababah et al. (2024) also analyzed millions of social media posts, along with newspaper and Fantasy Premier League data, for more implicit racial biases. Their work specifically focused on the treatment of minority players when they performed badly and found no pattern of such systemic biases. However, they do not control for hidden confounders that might impact fan discussions, such as a player's height, BMI, team, or position.

Measuring Racial Bias in Text. There is a large body of literature in natural language processing and social computing dedicated to measuring racial bias in textual data (Field et al. 2021). Racial bias has been studied through text analytics methods for auditing autocomplete (Ha, Kong, and Jhaver 2025), for characterizing hateful behavior online (Mittos et al. 2020; Chandrasekharan et al. 2022), and assessing racial stereotypes in digital content such as history books (Lucy et al. 2020), Wikipedia biographies (Field et al. 2022), and NLP tools (Sap et al. 2019; Field et al. 2023). Many of these methods attempt to detect overt racial bias, e.g., the use of racial slurs and stereotypes, while others assess differential behavior of NLP tools based on racialized dialects such as African-American English (Sap et al. 2019; Blodgett, Green, and O'Connor 2016). However, the conceptualization of race can be limited, especially to American constructs of race (Field et al. 2021). Studying racial biases with observational data also presents challenges in establishing causal effects: while there are some exceptions (Munger 2017), conducting experiments on social media to tease apart racial effects is ethically fraught, while there are questions of how to view perceived race as a 'treatment' when it is (usually) measured as an immutable variable (Hu and Kohler-Hausmann 2025).

In our work, we contextualize racial bias as an *implicit* measure, manifesting as differential treatment of Black Soccer players. Our work is also grounded in discussions of race/ethnicity in sports media, as discussed in the previous paragraphs. Therefore, we measure implicit racial bias as differences in sentiment, toxicity, and athletic stereotypes directed toward players of different racial backgrounds. Our setup is based on observational data, and we cannot establish causality; however, we control for several confounders that could be driving this type of differential treatment other than race to pinpoint its effect.

4 Data Collection

In the following, we describe the datasets that we collected for this study. In addition to social media data, we also needed to collect metadata on soccer players.

4.1 Social Media Data

Within the context of the English Premier League, FA Cup, and Champions League, we collected data from two social media platforms: Reddit and Twitter. This collection was taken in 2022, when both platforms still provided APIs that allowed scraping their data. By now, Twitter has also been rebranded as 𝕏, but we use the name of the platform at the time of data collection.

Reddit. One of the most popular online communities about soccer in the world is the `r/soccer` subreddit[5], which has close to 8 million subscribers and thousands of comments per day. This subreddit offers live and post-match threads for several soccer competitions, which are emphasized by the corresponding community bookmarks. For our analysis, we specifically scraped data from such match threads, as these are easy to identify and parse due to their standardized format containing the keywords 'Match Thread' or 'Post Match Thread', the names of the teams competing, and the name of the competition. We collected comments from all Premier League and FA Cup match threads, and from all match threads for Champions League games in which at least one Premier League team was competing during the 2017/18-2020/21 seasons. In this way, we collected more than 1,000,000 comments from Reddit.

Twitter. To complement the Reddit data, we scraped Twitter for tweets related to Premier League, FA Cup, and Champions League matches in the 2017/18-2020/2021 seasons. To do so, we used hashtags, as each Premier League team has some (mostly) unique hashtag(s) from which they could be identified[6]. Using this system, for each match in the given competitions and timeframe, we added a tweet to our data if it included a hashtag of these teams and if it had been posted within one calendar day before and after the match. In total, we collected almost 3,000,000 tweets.

4.2 Player Metadata

Despite the abundance of soccer-related websites and datasets, there was no single data source available that provided all the data needed for our analysis. Therefore, we ended up collecting data from the following three sources.

Transfermarkt. As our main source for player-related data, we scraped data from *Transfermarkt*,[7] a Germany-based online platform focusing on news and discussions about market values and transfers of soccer players. It also provides extensive information on players, teams, managers, and agents in international soccer. For each season, it gives an overview of all clubs that played in a major soccer league, and for each club, it provides a detailed overview of the players that have played for this club in a given season, including their positions, birthdates, height, strong foot, or their market values, but also, notably, the exact date that they joined the club. For our analysis, we scraped the season-wise player information of all Premier League teams.

WhoScored. Within our analyses, we also needed to control for player performance as a potential confounder. Toward that end, *WhoScored*[8] offers one of the most widely used

[5]https://www.reddit.com/r/soccer/

[6]https://www.thesun.co.uk/sport/football/4104194/premier-league-hashtags-on-twitter-arsenal-to-west-ham-united/

[7]https://www.transfermarkt.de/

[8]https://www.whoscored.com/

Comment	Name	Age	Race	Nationalities	Position	Club	Rating	Market Value	Virality	Sentiment	Toxicity	Supersenses
great **pass** by Lindelöf	Victor Lindelöf	25	White	Sweden	Central Defender	Manchester United	7.04	28M	610	0.62	0.01	Technical
why the f*** is that idiot jahanbaksh coming on	Alireza Jahanbakhsh	27	Other	**Iran**, Netherlands	Winger	Brighton & Hove Albion	6.47	5M	86	-0.54	0.87	-
TELLA IS **RAPID**	Nathan Tella	21	Black	Nigeria, England	Winger	Southampton	5.88	2M	35	0.0	0.14	Physical

(a) Snapshot of processed Reddit data

Comment	Name	Age	Race	Nationalities	Position	Club	Rating	Market Value	Virality	Sentiment	Toxicity	Supersenses
Best Defender in the World bar none ! The **Big** Man VVD #LFC #VVD	Virgil van Dijk	27	Black	Netherlands, Suriname	Central Defender	Liverpool	8.5	90M	3337	0.71	0.02	Physical
Declan Rice is so overhyped its unreal. He's fu***** bad! #WHUFC #WestHam	Declan Rice	20	White	England, Ireland	Central Midfielder	West Ham United	6.81	49.5M	2105	-0.49	0.89	-
Ghezzal is terrible should be nowhere near our team... shameful **selfishness** and loss of possession for that goal #lcf	Rachid Ghezzal	26	Other	**Algeria**, France	Winger	Leicester City	5.96	8M	203	-0.88	0.43	Mental

(b) Sample of processed Twitter data.

Table 1: *Snapshots of processed datasets*. **Green** words in comments indicate supersense terms which were identified in the comment. **Bold** face country names indicate a Muslim majority coutry. Blue highlights indicate rephrasing to preserve the privacy of authors as well as redaction of explicit terms.

performance ratings in the world of soccer, using data provided by *OPTA*[9] to calculate the ratings. WhoScored assigns ratings between 0 and 10, starting with 6.0 at the beginning of a game, which are updated every 30 seconds during the game. The calculation includes more than 200 raw statistics weighted according to the influence within the game. We exhaustively scraped match data for each individual Premier League, FA Cup, and Champions League match during the 2017/18-2020/21 seasons, including the performance ratings of all players involved in each fixture. Afterward, we merged the resulting data with the Transfermarkt data by first matching player names and then validating these matches by comparing the teams for which the matched players competed. A small number of players who have not been matched from the WhoScored data have been manually matched.

Soccer Wiki. While there are countless publicly available datasets on soccer players, few of these datasets contain information on their skin tones. An exception to this is *Soccer Wiki*[10], which is a collaborative database that anyone can create and edit data on and which contains information about soccer players, teams, managers, and leagues. Aside from skin color, the player profiles on Soccer Wiki also include information about each player's age, current club, and position (among others), as well as the URL of the player image on the webpage, which we also retrieved within this crawl.

An index of all active players included in their data can be downloaded directly from the page, and based on the player IDs provided in these data, one can easily navigate and scrape player profiles from this page. We matched the names of all players in the data with the list of all Premier League players obtained from Transfermarkt, limited to those that have been mentioned in the social media data, and then scraped the profiles of all Premier League players. Afterward, we validated the matched players by comparing the birthdates provided in both data sources. Where no exact matches of names were found due to differences between the datasets or due to retired players not being in the player index, we manually determined the IDs of these players to obtain one-to-one matches between these datasets.

[9]https://www.statsperform.com/opta/

[10]https://en.soccerwiki.org

Statistic	Reddit	Twitter
Number of comments	204,217	647,262
Number of players	1,065	1,251
Most viral player	Mohamed Salah	Harry Kane
Most viral club	Manchester United	Chelsea
Avg. sentiment score	0.043	0.169
Share of positive comments	25.11%	36.58%
Share negative comments	19.29%	13.77%
Avg. toxicity score	0.188	0.084
Share toxic comments	7.87%	1.65%
Share physical comments	1.99%	1.93%
Share technical comments	10.28%	9.52%
Share mental comments	3.51%	6.11%

Table 2: *Summary statistics of Reddit and Twitter data.*

5 Data Processing and Augmentation

Next, we describe the steps we took to process and augment our datasets. Sample snapshots of the processed datasets can be found in Table 1. An overview of statistics of the processed datasets can be found in Table 2. Further details about the datasets can be found in Appendix A.

5.1 Perceived Race of Players

In this study, our objective is to analyze racial biases in online discussions. Although determining people's race can be

	Black	Dark Brown	Brown	Olive	White	Pale
Black	48	194	132	5	0	0
Latino	0	1	50	73	12	0
Mid-East	0	0	4	42	20	0
Indian	0	0	14	6	0	0
East Asian	0	0	0	9	0	0
SE-Asian	0	0	1	1	0	0
White	0	0	5	80	547	11

Table 3: *Triangulation of perceived race of players*. Rows indicate the race predicted by FairFace, Columns indicate the skin color as provided by Soccer Wiki.

ethically and technically challenging, in the context of this paper, we hypothesize that a soccer player's *perceived* race affects the type of comments they receive on social media. We used a triangulation approach that combines the following two sources to determine the perceived race of a player.

1. **Soccer Wiki.** As discussed above, Soccer Wiki is a collaborative database of soccer players, which also contains information about their skin tones. Within this database, possible skin color values are black, brown, dark brown, olive, pale, and white.
2. **FairFace.** *FairFace* (Karkkainen and Joo 2021) is a dataset designed to be balanced with respect to racial representations. It considers a common race categorization as used by the US Census Bureau, which distinguishes between seven race categories: *White*, *Black*, *Indian*, *East Asian*, *Southeast Asian*, *Middle Eastern*, and *Latino*. Its creators have shown that models trained on this dataset give accurate and balanced predictions across different races, and thus we apply their model to player images retrieved from Soccer Wiki.

Table 3 shows a cross-tabulation of the corresponding values from these data sources. For our analysis, we consider a player as **White**, if their predicted race according to FairFace is *White*, and their skin color according to Soccer Wiki is *white* or *pale*. Similarly, we consider a player as **Black**, if their predicted race according to FairFace is *Black* and their skin color according to Soccer Wiki is *brown* or darker. In addition, 19 players did not have image or skin color information on Soccer Wiki. For these players, perceived race has been manually annotated. In total, of 1,273 players, 376 are classified as Black, 570 as White, and 327 as **Other**.

5.2 Identifying Player Names in Comments

To determine whether a player was mentioned in a comment, we applied a two-step approach. First, we performed named entity recognition and part-of-speech tagging on each comment to identify a set of candidate words or entities which could refer to soccer players. Second, we matched each candidate word with the names of the players who were involved in the game that was commented on.

For the first step, we applied the `spaCy` package and considered all named entities and nouns from a comment as candidates for player names. In addition, for the Twitter data, all tokens starting with the `'@'` character were considered an entity, given that these relate to account names. Although this procedure inherently produces a candidate set with many false positives, we found that more restrictive policies, such as only considering named entities, would leave large amounts of player names as false negatives.

The false positives from our candidate sets were filtered out in the second step, where we matched all nouns and entities with the names of players who participated in the match that was commented on. For player names, we considered their short names, that is, last names or aliases such as *Rodri*, their full names, and, for players with compound last names such as Kevin de Bruyne or Trent Alexander-Arnold, also their initials, since acronyms like "KdB" or "TAA" are commonly used to refer to these players. The matches were then determined using fuzzy string matching techniques based on Levenshtein distance, as implemented in the `rapidfuzz` package. Specifically, we considered a candidate word a match with a player name if

1. a candidate word exactly matches short name or initials,
2. a candidate word and short name are longer than five characters and have a similarity ratio higher than 80,
3. a candidate word and long name have a token set similarity ratio higher than 80,
4. a candidate word starts with `'@'` (i.e., corresponds to a Twitter account handle) and contains the short name of a player as substring, or
5. a candidate word was identified a named entity, has at least four characters, and the short name starts with it.

Criterion 5 was designed to match players with longer last names, such as Pierre-Emerick Aubameyang, for which abbreviations such as "Auba" are commonly used. Except for criterion 4, candidate words and player names were always normalized to title case. These criteria have been optimized iteratively with the aim of including all common ways players are referred to, while also minimizing false positives.

For every comment, we recorded every match with player names. However, for our analysis, we only considered comments in which exactly one single player has been mentioned. For Reddit, we thereby filtered out 36,538 comments (15%), for Twitter, 251,946 comments were removed (28%).

5.3 Augmentation of Comments

We conceptualize racial bias towards non-White soccer players as an implicit measure rather than explicit racial slurs. In principle, explicit racial content is trivial to detect using a list of racial slurs. However, we are interested in more subtle differences in the treatment of players. Therefore, we investigate three dimensions of differential treatment: sentiment towards a player, toxicity towards a player, and supersenses of comments about a player. In the following, we describe how we augmented the social media comments along these three dimensions.

Sentiment. We used VADER (Hutto and Gilbert 2014) to identify the sentiment of tweets and Reddit content because it is specifically built to capture sentiment expressed in social media contexts and has been shown to work well in that domain (Guimaraes et al. 2019; Joseph et al. 2019; Monselise and Yang 2022). It takes into account excessive punctuation and common slang, and it not only reports on the positivity and negativity score, but also tells us how positive or negative a feeling is (Hutto and Gilbert 2014). The compound measure, which is a value between -1 and 1, indicates whether the text is positive or negative. We consider comments with a value smaller than -0.4 to be of negative sentiment, comments with a value greater than 0.4 positive, and the remaining comments neutral.

Toxicity. While negative sentiment can be indicative of racial bias, a more precise and complementary variable is offensive, hateful, or abusive language directed at darker-skinned players. Therefore, to model this type of language, we used the Perspective API toxicity endpoint (Lees et al. 2022). This API is widely used in computational social science studies (Rajadesingan, Resnick, and Budak 2020; Hua, Ristenpart, and Naaman 2020). However, researchers have pointed out racial biases within the Perspective API, specifically the tendency to label content with African American dialect as more toxic (Sap et al. 2019). To ensure that our results are not affected by similar biases, two authors of this paper validated the results of Perspective API on a small sample (100 comments each from Twitter and Reddit). Human labels achieve a high overlap with Perspective's labels (88% for Reddit and 92% for Twitter). Although automated toxicity labels are not perfect, we consider this level of match to be sufficient for downstream analysis. The false positives from Perspective are mainly due to the presence of profanity, which was, however, not directed towards a particular player (e.g., "Holy shit, that was a solid save!").

Soccer Supersenses. Negative comments alone are not necessarily indicative of racial bias, and racial stereotypes can become evident through the presence of specific messages in different contexts. As previous research pointed out, Black players often received more praise related to their physical skills and more criticism regarding their technical or tactical abilities. To analyze whether such disparities also exist on Reddit or Twitter, we applied a dictionary-based approach to classify whether a given post contains terms related to **soccer supersenses**, i.e., one of the following three categories.

- **Physical** terms relate to the athleticism of players and mainly concern their pace (e.g., *quick, fast, agile)* or physical strength (e.g., *athletic, strong, bulky*).
- **Technical** terms describe nonphysical abilities, and include terms relating to ball-handling actions on the field, such as *dribbling, passing, shooting*, as well as adjectives describing player skills, such as *technical, skillful*.
- **Mental** terms more broadly relate to intelligence (e.g., *clever, creative, mindless*), or character traits (e.g., *disciplined, aggressive, selfish*) of a player.

The corresponding dictionary, which assigns a list of terms to these categories, has been assembled manually, using existing categorizations of soccer terms (Campbell and Bebb 2021; McLoughlin 2023) as a foundation. Building on these, additional terms have been added based on domain knowledge, and a thesaurus has been used to exhaustively include as many synonyms as possible. In this way, the resulting dictionary contains more than 400 terms. Because we parsed posts in their original form without stemming tokens, these, however, include several variants of the same terms, such as verbs in different tenses. Using this resulting dictionary, we parsed every post from our datasets for occurrence of these terms and documented the occurrence of these terms along with the corresponding category.

For several terms that have multiple word senses, such as *big*, *strong*, or *weak*, we found, however, that these are often used in contexts not related to the intended supersenses, such as simply characterizing a player's performance as *strong*. Therefore, for these ambiguous terms, we further post-processed comments containing such terms via a large language model (LLM), specifically `Llama-3-8B` (Meta AI 2024). We prompted the LLM to identify whether comments using such terms are indeed related to the sought-after soccer supersenses. The 51 words that we disambiguated, as well as the prompts used for disambiguating their specific word senses, are included in Appendix B. We provide the full dictionary of terms in our online repository.

6 Methods

To answer the research question introduced in Section 1, i.e., to assess the relationship between a player's race and the type of social media commentary targeted towards them, we conduct (i) a regression analysis, and (ii) a matching-based analysis. In both analyses, we specifically investigate differences in terms of a) sentiment, b) toxic content, and c) type and valence of soccer supersenses that are commented on, as described in Section 5.3. In the following, we provide more details about the regression and matching approaches.

Regression Analysis. We conduct simple linear regression analyses in which we predict average sentiment scores, toxicity scores, and the share of comments with physical, mental, or technical supersenses on player-level. The main independent variable we are interested in is race, which we model as a categorical variable with *Black* as the reference category. In addition, we consider the following attributes, which were obtained from the comments and player metadata.

- **Age.** Since young players are naturally met with different expectations than more experienced players, we control for the age of a player via their year of birth.
- **Height.** Short players are often expected to have better technical skills, while taller players are often perceived to be physically superior. Thus, we include height in our analysis, which was computed from the average height values given by Transfermarkt and Soccer Wiki.
- **Body Mass Index (BMI).** Since we also analyze comments about physical attributes of players, we need to control for their physical bulk. Toward that end, we consider their BMI, which we compute from the weight data provided by Soccer Wiki.
- **Position.** A player's position may also influence the sentiment of the comments they receive, as, for instance, attacking players who are more likely to score goals might, in turn, also receive more praise. We therefore include each player's position in our analyses but group the positions provided in the Transfermarkt data into the following seven groups: keeper, central defender, wide defender, central midfielder, attacking midfielder, winger,

and forward. That way, we reduced natural fuzziness in finer positions while still distinguishing between central and peripheral positions, in which players with darker skin tones tend to be overrepresented (Nobis and Lazaridou 2023). Position is modeled as a categorical variable in our regressions.

- **Big Six Clubs.** Clubs differ in their worldwide popularity, and the popularity of a club can affect the sentiment towards their players. To account for this, we add an attribute indicating whether a player was active for one of the *big six* (Conn 2021) most popular and competitive clubs in Premier League during the seasons under study, namely *Arsenal*, *Chelsea*, *Tottenham*, *Liverpool*, *Manchester City*, and *Manchester United*.
- **Virality.** High popularity of individual players may lead to more polarized commentary about them. Thus, we consider each player's virality for our analysis. This is measured in terms of the total number of comments received on each platform, of which we then took the logarithm due to the nonlinear distribution of this variable.
- **Performance.** The sentiment of comments about players could be influenced by their performance. Thus, we consider the average performance rating of a player across the four seasons under study for our analysis.
- **Market Value.** Another indicator of the quality of players, and implicitly also the expectations associated with them, is their market value. In our regression, we consider the maximum market value of each player during the four seasons under consideration and used its logarithm due to the exponential nature of these values.
- **Continent.** The nationality of a player may also cause some biases associated with them. As including more than 100 countries in our analysis may cause overfitting, we consider the corresponding continents. This is done by one-hot encoding, thereby allowing for multiple continents if a player has two nationalities.
- **Muslim Country.** Aside from race and nationality, there may also be biases toward players based on religion. Since we do not have data about the religion of players, we use their nationalities as a proxy and consider them to be from a Muslim country if one nationality was a country with a Muslim majority population according to Hackett et al. (2025).

We fit our regression models via ordinary least squares, without interactions of independent variables. To ensure consistency across the regression weights, we normalize all non-categorical independent variables to the interval $[0, 1]$. For the soccer supersense regressions, we exclude players with fewer than 20 comments—given that less than 5% of comments contained corresponding terms, these entries could otherwise bias our results.

Matching Analysis. Our matching analysis is also conducted on player level, where we compare a) average sentiment score, b) average toxicity score, and c) fraction of comments with physical, mental, and technical supersense between Black players and White players. Players of other races are excluded to avoid ambiguities.

To form the groups of players to compare, we consider the age, height, BMI, position, virality, market value, and big six club attributes, which we also used in our regression. Based on these attributes, we apply a greedy 1-1 matching approach, where for each Black player, we first restrict the candidates to players of the same position. Afterward, we select the closest match based on Mahalanobis distance with respect to the remaining player attributes. After inspecting the matches we initially obtained, we decided to only match pairs of players with a Mahalanobis distance $d < 1.75$. We use the same matched groups across both datasets, since with the exception of virality—which we aggregate as the logarithmic sum across both databases—all attributes were dataset-agnostic. To avoid outliers, we only consider players in our matching who had at least 20 comments in both datasets. In the end, two groups of exactly 100 players are matched—in Appendix D, we provide the full list of all matches. Based on these matched groups, we compute the aggregated target variables and evaluate their difference using a paired sample t-test.

Multiple Hypothesis Testing. In our analyses, we essentially evaluate multiple hypotheses on fixed datasets. Thus, we need to adjust for the multiple comparison problem. Since each hypothesis is evaluated on two independent datasets, we account for this in the following way. First, we report the default p-values resulting from each combination of target variable and dataset. These p-values are then combined on dataset level via Fisher's method before applying the Bonferroni correction on the resulting p-values. That way, we can reliably report whether we have a statistically significant effect on at least one dataset, while avoiding overly conservative corrections for multiple comparisons.

7 Findings

The results of the regression analyses with respect to sentiment, toxicity, and the physical supersense are summarized in Figure 1, and the main results from our matching analysis are shown in Table 4. Results for the remaining supersenses (Figure 3, along with additional details on the results and further regression variants, can be found in Appendix C. For the regressions, we generally observe that Twitter models have a better overall fit than Reddit: only for toxicity, Reddit yields a higher R^2 value. At the same time, the R^2 values are generally low, ranging between 0.07-0.37.

In the following, we discuss how the primary independent variable of interest, the race of a player, is associated with the sentiment of comments they receive, the toxicity of comments, and the types of supersense-related comments.

RQ1: Sentiment. For both platforms, we observe that positive sentiment is positively correlated with a player being White, however, on none of these platforms, this effect is significant. Similarly, for players of Other perceived race, we do not find a significant positive correlation with sentiment on either platform. Unexpectedly, players with higher market value receive less positive sentiment, potentially reflecting higher expectations and criticism. Finally, we also see an age bias on both platforms, with older players being less likely to receive positive comments. For both platforms, player age is significantly correlated with sentiment——older

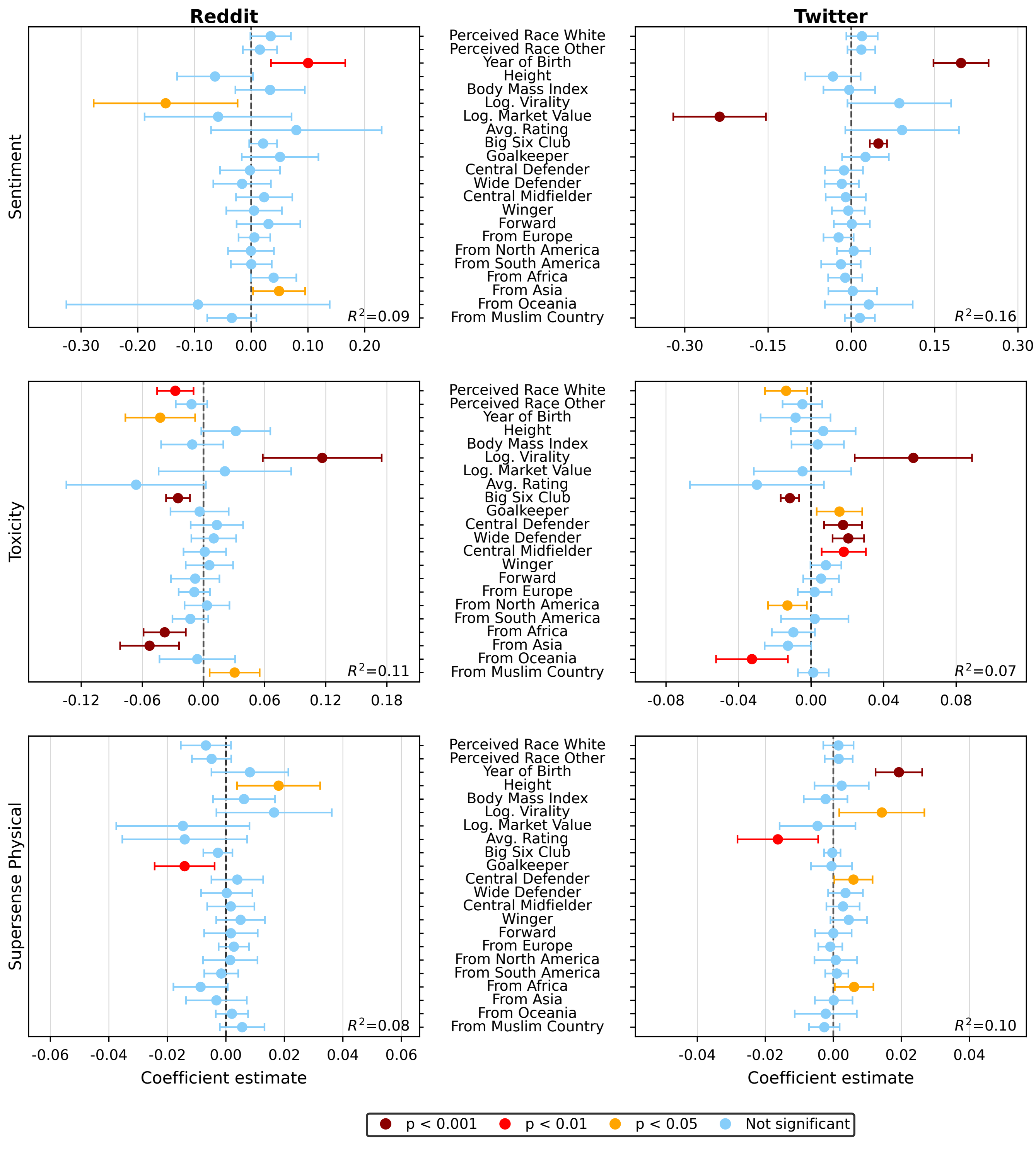

Figure 1: *Regression results.* Each panel depicts coefficient estimates of independent variables, along with 95% confidence intervals, from the regressions on each combination of dependent variable (panel rows) and dataset (panel colums). Results for the *Mental* and *Technical* supersenses are shown in Appendix C.

	Reddit				Twitter				Combined p-value	
Target Variable	Black	White	Diff	p-value	Black	White	Diff	p-value	Fisher	Bonferroni
Average Sentiment Score	0.0473	0.0656	-0.0183	0.0378	0.1537	0.1550	-0.0012	0.8976	0.1487	0.7435
Average Toxicity Score	0.1850	0.1749	0.0101	0.0829	0.0900	0.0857	0.0043	0.2834	0.1116	0.5579
Share Supersense Physical	**0.0261**	**0.0199**	**0.0063**	**0.0514**	**0.0247**	**0.0198**	**0.0049**	**0.0040**	**0.0019**	**0.0097**
Share Supersense Technical	0.0885	0.1005	-0.0120	0.0310	0.0933	0.0980	-0.0048	0.3614	0.0614	0.3072
Share Supersense Mental	0.0367	0.0342	0.0026	0.3809	0.0588	0.0605	-0.0018	0.5275	0.5234	1.0000
Share Supersense Physical (+)	**0.0312**	**0.0231**	**0.0081**	**0.0709**	**0.0361**	**0.0256**	**0.0105**	**0.0004**	**0.0003**	**0.0010**
Share Supersense Technical (+)	0.1192	0.1323	-0.0131	0.1955	0.1146	0.1201	-0.0055	0.4418	0.2979	0.8938
Share Supersense Mental (+)	0.0478	0.0473	0.0004	0.9443	0.0781	0.0825	-0.0043	0.3025	0.6436	1.0000

Table 4: *Matching results.* We show averaged sentiment and toxicity scores along with aggregated shares of comments relating to the three soccer supersenses, separated by perceived race. Lower three rows consider supersenses aggregated over comments with positive sentiment, i.e., comments in which players were praised. We report dataset-specific p-values resulting from a paired sample test, as well as the p-value resulting from a combining dataset-specific p-values via Fisher's method, and then applying the Bonferroni correction. Bold face rows indicate statistically significant differences.

players receive more negative sentiment. Further, positive sentiment on Twitter is negatively correlated with virality. Our matching analysis generally corroborates these findings, as we also observe that average sentiment appears to be lower for Black players across both platforms; however, for none of these platforms this effect is statistically significant.

RQ2: Toxicity. From Table 2, we see that the Perspective API generally labels Reddit comments as more toxic than tweets. This could indicate differences in moderation between the two platforms, and, in turn, such higher deletion rates may have skewed the results. Ultimately, we do find a significant correlation between race and the toxicity of comments on both platforms, with White players obtaining less toxic comments than Black players. On Reddit, this effect seems even more pronounced ($\beta = -0.028, p = 0.0024$). After combining the p-values of both datasets using Fisher's method and applying the Bonferroni correction, we still obtain a p-value < 0.01, indicating that in at least one dataset there is indeed a significant effect. Next to race, we note that for both platforms, players from popular clubs are associated with less toxicity (Twitter: $\beta = -0.01, p < 0.001$, Reddit: $\beta = -0.024, p < 0.001$). Further, on Twitter we see that defenders generally tend to get significantly more toxic comments, whereas on Reddit, we observe negative correlations between toxicity and players being from Asia or Africa. In Appendix C, we include results on associations between race and other toxic commentary, such as insults and profanity.

In our matching analysis, we also observe higher toxicity for the Black players, and while on Reddit, this effect seems quite pronounced, the combined and corrected p-value across both datasets does not allow for rejection of the null hypothesis that on none of the datasets, Black and White players receive comments of the same toxicity.

RQ3: Discussions of Soccer Supersenses. Our final analyses focus on the semantic content of language used—whether users emphasize technical ability (positive or negative), physical traits, or mental attributes of players. In our regressions, we overall do not find any significant correlations between perceived race and supersenses. By contrast, our matching analysis shows some effects that are consistent with previous literature (Campbell and Bebb 2021; McLoughlin 2023). Specifically, on both platforms, we observe that Black players receive more comments related to their physical attributes and fewer comments related to their technical abilities than White players. While for the technical supersense, we cannot completely reject the null hypothesis that both groups receive the same amount of comments regarding their technical skills, we obtain strong statistical significance ($p < 0.01$) for the physical supersense, indicating that at least in one dataset the observed effect is not random.

8 Discussion

We now contextualize our findings with the research questions we set out to answer, implications for the design of social media platforms, and limitations of our work, as well as ethical considerations.

Summary of Findings. Regarding our main research questions, we find that White players receive slightly fewer toxic comments (RQ2) on both Twitter and Reddit. This is specifically evident in the regression analysis, with the matching analysis indicating a similar trend. Using the matching analysis, we also observe a racialized attribution of abilities (RQ3): Black players receive more comments regarding their physical attributes, which appears particularly pronounced on Twitter. Conversely, especially on Reddit, we find that White players receive more comments regarding their technical qualities, although we cannot establish statistical significance after correcting for multiple hypotheses. Yet, these patterns mirror long-standing concerns in sports media studies, where Black athletes are often framed as physically superior but mentally or technically inferior (Campbell and Bebb 2021; McLoughlin 2023). We also find a negative result: positive sentiment does *not* seem to be associated with player race to a degree that is statistically significant (RQ1). This appears more in line with the findings of Alrababah et al. (2024).

We note that across both datasets, the explained variance in our regressions is relatively modest, indicating that the sentiment, toxicity level, and soccer supersenses invoked in social media comments are affected by unobserved contextual or behavioral factors beyond a player's race or soccer-related characteristics. Yet, these predictors still provide meaningful insight, particularly due to the consistency of their directionality. Our findings provide evidence that race, player position, and performance metrics significantly influence the tone and toxicity of online discourse about soccer players, but we also uncover differences across platforms—Twitter sentiment is more sensitive to a player's club, while toxicity on Reddit is more influenced by player age.

Implications for Platform Design. Our results suggest several design implications for social media platforms aiming to address and mitigate racial bias. Platform moderation tools often focus on detecting explicit hate speech. However, we find that there are still toxic comments towards players, especially on Reddit. Such types of comments are more likely to target Black soccer players across both platforms. Furthermore, our matching-based analysis shows that racial bias can also be subtle and structural, embedded in how the language around players is framed. Platforms should incorporate context-aware monitoring that can capture these nuances. While we do not advocate for censorship of implicit racial abuse, it might still be good to make audiences and the authors of the posts themselves aware of the nature and potential consequences of such content. By providing open-access, soccer-specific datasets and dictionaries of stereotype-laden terms, we encourage the development of models that can recognize racial stereotyping even when toxicity is absent. Just as platforms experiment with features like comment filters for creators, athletes, and public figures, especially those who do not have social media teams, might benefit from opt-in tools that suppress or summarize commentary likely to follow biased patterns (Jhaver et al. 2022).

Limitations. One major limitation of this study is that the overall variance explained (adjusted R^2) in the regression analysis is low for many dependent variables (< 0.1). However, small effect sizes are often typical for predicting human behavior and can still reflect meaningful effects in complex social phenomena (Abelson 1985; Rosenthal and Rubin 1979). Nonetheless, many of the independent variables do have significant correlations, indicating that our models are informative. We do not advocate using our models to predict the behavior of commentors on social media, but rather to unearth potential factors associated with their opinions. Next, we only had access to social media posts that have not been deleted, moderated, or removed from the platforms. This naturally affects some of the measures, particularly toxicity. It would be interesting to conduct a similar study with live-streamed data to estimate the proportion of explicitly racist and hateful comments that are posted during the match and taken offline afterward due to platform moderation. There may also be missing data from Twitter since only comments that used specific hashtags were extracted. Further, for each hashtag, there was a limit of 2,000 tweets that could be retrieved, which could affect the final results.

In addition, we only considered comments in which a single player is mentioned, which meant that around 15% of Reddit data and 28% of Twitter data was excluded from our analysis. This was done to avoid attributing the sentiment of a comment toward the wrong player; however, in some cases, a comment may still not have been addressed at the player that has been identified in this comment, but at other people such as coaches or other platforms users. Future studies could analyze comments at sentence level and explore more advanced methods in determining which entity comments at sentence levels are directed to.

Finally, we used a variety of automated techniques to augment social media comments, for which perfect accuracy cannot be expected. However, we put a lot of thought into maximizing the correctness of our approach. We obtained the sentiment of social media comments using VADER, which has been validated for social media analysis and widely used (Hanley, Kumar, and Durumeric 2023). To obtain the toxicity of a comment, we used the Perspective API and conducted validations of its results to ensure robust findings (cf. Section 5.3). Perspective API has been shown to have racial biases of its own (Sap et al. 2019), where it rates African American English (AAE) as more toxic; however, our dataset does not focus on US-based discussions, so the chances of misclassified AAE are low: we also did not find any occurrence of it during our manual validation. Lastly, we use a soccer-specific dictionary that labels fine-grained dimensions of players' soccer supersense, whose precision we further hone through an LLM-based disambiguation step.

Ethical Considerations. Race is a complex feature that consists of several constitutive parts (Sen and Wasow 2016). We used multiple sources to triangulate the *perceived race* of soccer players, rather than the self-identified race. Since we are interested in how players are perceived by social media commentors, we believe this is a more useful proxy. To construct this perceived racial category, we use a combination of automated and manual analysis—the crowdsourced labels from Soccer Wiki and the FairFace model. While our approach is based on methodological triangulation and human validation, making it more robust than simply relying on automatic computer vision, our final categorization uses three categories: White, Black, and Other. We acknowledge that these categories do not capture the full spectrum of implicit racial abuse a player might experience.

9 Conclusion

Our study provides a large-scale multiplatform investigation of racial biases in online soccer discourse. Using comments from Twitter and Reddit about Premier League players over four seasons, we explored the relationship between a player's perceived race and the sentiment, toxicity, and content of online discussions. We found that White players tend to receive less toxic comments than Black players. In addition, we uncovered patterns of implicit racial stereotyping in how technical, mental, and physical abilities of players are discussed, with Black players receiving more comments related to their physicality. Future research could conduct similar analyses on different platforms or investigate other kinds of bias, such as differences in discussions about men's and women's soccer.

Acknowledgements

The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1597-1 FUGG.

References

Abelson, R. P. 1985. A Variance Explanation Paradox: When a Little Is a Lot. *Psychological Bulletin*, 97(1): 129–133.

Alrababah, A.; Marble, W.; Mousa, S.; and Siegel, A. A. 2024. Are Minorities Punished More Harshly for Underperformance? Evidence from Premier League Soccer. OSF Preprints:7d2cu_v1.

Bennet, H.; and Jönsson, A. 2017. Klick It out: Tackling Online Discrimination in Football. In Kilvington, D.; and Price, J., eds., *Sport and Discrimination*, number 72 in Routledge Research in Sport, Culture and Society, 215–226. Milton Park, United Kingdom: Routledge.

Bliuc, A.-M.; Faulkner, N.; Jakubowicz, A.; and McGarty, C. 2018. Online Networks of Racial Hate: A Systematic Review of 10 years of Research on Cyber-Racism. *Computers in Human Behavior*, 87: 75–86.

Blodgett, S. L.; Green, L.; and O'Connor, B. 2016. Demographic Dialectal Variation in Social Media: A Case Study of African-American English. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1119–1130. Austin, TX, United States.

Brewin, J. 2024. Emotional Vinícius Júnior Losing Desire to Play after Racist Abuse in Spain. *The Guardian*. https://www.theguardian.com/football/2024/mar/25/vinicius-junior-racist-abuse-spain-brazil-football-real-madrid. Accessed: 2026-04-03.

Cable, J.; Kilvington, D.; and Mottershead, G. 2022. 'Racist Behaviour Is Interfering with the Game': Exploring Football Fans' Online Responses to Accusations of Racism in Football. *Soccer & Society*, 23(8): 880–893.

Campbell, P. I.; and Bebb, L. 2021. 'He Is like a Gazelle (When He Runs)' (Re)Constructing Race and Nation in Match-Day Commentary at the Men's 2018 FIFA World Cup. *Sport in Society*, 25(1): 144–162.

Castaño-Pulgarín, S. A.; Suárez-Betancur, N.; Vega, L. M. T.; and López, H. M. H. 2021. Internet, Social Media and Online Hate Speech. Systematic Review. *Aggression and Violent Behavior*, 58: 101608.

Chandrasekharan, E.; Jhaver, S.; Bruckman, A.; and Gilbert, E. 2022. Quarantined! Examining the Effects of a Community-Wide Moderation Intervention on Reddit. *ACM Transactions on Computer-Human Interaction*, 29(4): 29.

Cleland, J. 2014. Racism, Football Fans, and Online Message Boards: How Social Media Has Added a New Dimension to Racist Discourse in English Football. *Journal of Sport and Social Issues*, 38(5): 415–431.

Conn, D. 2021. Premier League's Big Six and Why They Want to Join a European Super League. *The Guardian*. https://www.theguardian.com/football/2021/apr/19/super-league-plan-english-clubs-european-football. Accessed: 2026-04-03.

Cullen, A.; and Williams, M. 2023. Online Hate Speech Targeting the England and Wales Men's Football Teams during the 2022 FIFA World Cup. Technical report, HateLab, Cardiff University, Cardiff, United Kingdom.

Field, A.; Blodgett, S. L.; Waseem, Z.; and Tsvetkov, Y. 2021. A Survey of Race, Racism, and Anti-Racism in NLP. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1905–1925. Virtual Conference.

Field, A.; Coston, A.; Gandhi, N.; Chouldechova, A.; Putnam-Hornstein, E.; Steier, D.; and Tsvetkov, Y. 2023. Examining Risks of Racial Biases in NLP Tools for Child Protective Services. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1479–1492. Chicago, IL, United States.

Field, A.; Park, C. Y.; Lin, K. Z.; and Tsvetkov, Y. 2022. Controlled Analyses of Social Biases in Wikipedia Bios. In *Proceedings of the ACM Web Conference 2022*, 2624–2635. Lyon, France.

FIFA. 2007. FIFA Big Count 2006: 270 Million People Active in Football. https://digitalhub.fifa.com/m/55621f9fdc8ea7b4/original/mzid0qmguixkcmruvema-pdf.pdf. Accessed: 2026-04-03.

FORCE11. 2020. The FAIR Data Principles. https://doi.org/10.18908/a.2019112601. Accessed: 2026-04-03.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Daumé III, H.; and Crawford, K. 2021. Datasheets for Datasets. *Communications of the ACM*, 64(12): 86–92.

Glynn, E.; and Brown, D. H. K. 2023. Discrimination on Football Twitter: The Role of Humour in the Othering of Minorities. *Sport in Society*, 26(8): 1432–1454.

Guimaraes, A.; Balalau, O.; Terolli, E.; and Weikum, G. 2019. Analyzing the Traits and Anomalies of Political Discussions on Reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 205–213. Palo Alto, CA, United States.

Ha, E.; Kong, H.; and Jhaver, S. 2025. Examining Racial Stereotypes in YouTube Autocomplete Suggestions. *New Media & Society*.

Hackett, C.; Stonawski, M.; Tong, Y.; Kramer, S.; Shi, A.; and Fahmy, D. 2025. The World's Religious Groups: How Their Sizes Changed from 2010 to 2020. https://www.pewresearch.org/religion/2025/06/09/how-the-global-religious-landscape-changed-from-2010-to-2020/. Accessed: 2026-04-03.

Hanley, H. W. A.; Kumar, D.; and Durumeric, Z. 2023. "A Special Operation": A Quantitative Approach to Dissecting and Comparing Different Media Ecosystems' Coverage of the Russo-Ukrainian War. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 339–350. Limassol, Cyprus.

Hu, L.; and Kohler-Hausmann, I. 2025. What Is Perceived When Race Is Perceived and Why It Matters for Causal Inference and Discrimination Studies. *Law & Society Review*, 59(2): 239–264.

Hua, Y.; Ristenpart, T.; and Naaman, M. 2020. Towards Measuring Adversarial Twitter Interactions against Candidates in the US Midterm Elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 272–282. Virtual Conference.

Hutto, C.; and Gilbert, E. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 216–225. Ann Arbor, MI, United States.

Jhaver, S.; Chen, Q. Z.; Knauss, D.; and Zhang, A. X. 2022. Designing Word Filter Tools for Creator-led Comment Moderation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 205. New Orleans, LA, United States.

Joseph, K.; Swire-Thompson, B.; Masuga, H.; Baum, M. A.; and Lazer, D. 2019. Polarized, Together: Comparing Partisan Support for Trump's Tweets Using Survey and Platform-Based Measures. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 290–301. Munich, Germany.

Karkkainen, K.; and Joo, J. 2021. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In *2021 IEEE Winter Conference on Applications of Computer Vision*, 1548–1558. Virtual Conference.

Kick It Out. 2025. Incident Reporting 2024/25. https://www.kickitout.org/sites/default/files/2025-08/Reporting%20-%20End-of-Season%20Stats%2024-25.pdf. Accessed: 2026-04-03.

Lees, A.; Tran, V. Q.; Tay, Y.; Sorensen, J.; Gupta, J.; Metzler, D.; and Vasserman, L. 2022. A New Generation of Perspective API: Efficient Multilingual Character-level Transformers. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3197–3207. Washington, DC, United States.

Longas Luque, C.; Sommier, M.; and van Sterkenburg, J. 2024. Talking Football: Discourses about Race/Ethnicity among Spanish Youth. *European Journal of Cultural Studies*, 27(5): 836–855.

Longas Luque, C.; and van Sterkenburg, J. 2022a. Constructing 'Race/Ethnicity' and Nationality in Spanish Media: A Content Analysis of International Football Coverage. *Media, Culture & Society*, 44(6): 1182–1198.

Longas Luque, C.; and van Sterkenburg, J. 2022b. Exploring Discourses About Race/Ethnicity in a Spanish TV Football Program. *Communication & Sport*, 10(6): 1113–1133.

Lucy, L.; Demszky, D.; Bromley, P.; and Jurafsky, D. 2020. Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks. *AERA Open*, 6(3).

Lukiv, J. 2024. German Football Coach Condemns Broadcaster's 'racist' Survey. *BBC News*. https://www.bbc.com/news/articles/cw4437m0gkpo. Accessed: 2026-04-03.

Magistro, B.; and Wack, M. 2023. Racial Bias in Fans and Officials: Evidence from the Italian Serie A. *Sociology*, 57(6): 1302–1321.

Matamoros-Fernández, A.; and Farkas, J. 2021. Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television & New Media*, 22(2): 205–224.

McCarthy, D.; Jones, R. L.; and Potrac, P. 2003. Constructing Images and Interpreting Realities: The Case of the Black Soccer Player on Television. *International Review for the Sociology of Sport*, 38(2): 217–238.

McLoughlin, D. 2023. Racial Bias in Football Commentary (Study): The Pace and Power Effect. Technical report, RunRepeat, Frederiksberg, Denmark.

Meta AI. 2024. The Llama 3 Herd of Models. arXiv:2407.21783.

Miranda, S.; Gouveia, C.; Di Fátima, B.; and Antunes, A. C. 2024. Hate Speech on Social Media: Behaviour of Portuguese Football Fans on Facebook. *Soccer & Society*, 25(1): 76–91.

Mittos, A.; Zannettou, S.; Blackburn, J.; and Cristofaro, E. D. 2020. "And We Will Fight for Our Race!" A Measurement Study of Genetic Testing Conversations on Reddit and 4chan. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 452–463. Virtual Conference.

Monselise, M.; and Yang, C. C. 2022. "I'm Always in so Much Pain and No One Will Understand" - Detecting Patterns in Suicidal Ideation on Reddit. In *Companion Proceedings of the Web Conference 2022*, 686–691. Lyon, France.

Moore, C. 2024. Premier League's Jaw-Dropping Viewership Crushes NFL. https://worldsoccertalk.com/news/premier-leagues-jaw-dropping-viewership-crushes-nfl/. Accessed: 2026-04-03.

Müller, F.; van Zoonen, L.; and de Roode, L. 2007. Accidental Racists: Experiences and Contradictions of Racism in Local Amsterdam Soccer Fan Culture. *Soccer & Society*, 8(2-3): 335–350.

Munger, K. 2017. Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Political Behavior*, 39(3): 629–649.

Nobis, T.; and Lazaridou, F. 2023. Racist Stacking in Professional Soccer in Germany. *International Review for the Sociology of Sport*, 58(1): 23–42.

Principe, F.; and van Ours, J. C. 2022. Racial Bias in Newspaper Ratings of Professional Football Players. *European Economic Review*, 141: 103980.

Rajadesingan, A.; Resnick, P.; and Budak, C. 2020. Quick, Community-Specific Learning: How Distinctive Toxicity Norms Are Maintained in Political Subreddits. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 557–568. Virtual Conference.

Richter, F. 2025. Super Bowl Pales in Comparison to the Biggest Game in Soccer. https://www.statista.com/chart/16875/super-bowl-viewership-vs-world-cup-final. Accessed: 2026-04-03.

Rosenthal, R.; and Rubin, D. B. 1979. A Note on Percent Variance Explained as A Measure of the Importance of Effects. *Journal of Applied Social Psychology*, 9(5): 395–396.

Sap, M.; Card, D.; Gabriel, S.; Choi, Y.; and Smith, N. A. 2019. The Risk of Racial Bias in Hate Speech Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–1678. Florence, Italy.

Sen, M.; and Wasow, O. 2016. Race as a Bundle of Sticks: Designs That Estimate Effects of Seemingly Immutable Characteristics. *Annual Review of Political Science*, 19: 499–522.

van Campenhout, G.; van Lienden, A.; and van Sterkenburg, J. 2023. Meanings given to (Super-)Diversity in the Dutch National Team by Dutch Football Commentators: A Historical Approach. *International Review for the Sociology of Sport*, 58(4): 647–665.

van Lienden, A.; and van Sterkenburg, J. 2023. Representations of Race/Ethnicity and the Nation: A Content Analysis of Televised Polish International Football. *International Review for the Sociology of Sport*, 58(1): 3–22.

van Lienden, A.; van Sterkenburg, J.; and Sommier, M. 2023. Meanings given to Race/Ethnicity in Everyday Football Talk by Young Adult Polish Audiences: A Reception Study. *Journal of Multicultural Discourses*, 18(3): 175–190.

van Sterkenburg, J.; Knoppers, A.; and de Leeuw, S. 2012. Constructing Racial/Ethnic Difference in and Through Dutch Televised Soccer Commentary. *Journal of Sport and Social Issues*, 36(4): 422–442.

van Sterkenburg, J.; Peeters, R.; and van Amsterdam, N. 2019. Everyday Racism and Constructions of Racial/Ethnic Difference in and through Football Talk. *European Journal of Cultural Studies*, 22(2): 195–212.

Vidgen, B.; Chung, Y.-L.; Johansson, P.; Kirk, H. R.; Williams, A.; Hale, S. A.; Margetts, H. Z.; Röttger, P.; and Sprejer, L. 2022. Tracking Abuse on Twitter Against Football Players in the 2021 – 22 Premier League Season. SSRN:4403913.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? No
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes
   (g) Did you discuss any potential misuse of your work? N/A
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we plan to make our dataset available after removing all personally identifiable information and through a gated access so that the data can be reused for research purposes only. For the peer review process we provide the code and aggregated data in an anonymized repository.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes
   (b) Have you provided justifications for all theoretical results? Yes
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? N/a
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, we control for several confounders
   (e) Did you address potential biases or limitations in your theoretical framework? Yes
   (f) Have you related your theoretical results to the existing literature in social science? Yes, see "Related Work"
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes
3. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? N/a, we do not train any classifiers but use opnely available off-the-shelf ones.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes, we also conduct thorough manual validations of all automated classifications.
4. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes, e.g., VADER (Hutto and Gilbert 2014)
   (b) Did you mention the license of the assets? N/a
   (c) Did you include any new assets in the supplemental material or as a URL? Yes
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? N/a. We use openly available social media data. While this does not mean that the data subjects have consented to their data being used for our observational study, the topic of it, i.e., racism towards soccer players on social media, is a societally important topic that we believe justifies the use of social media data. Furthermore, when releasing the data to ensure reproducibility, we have removed all personally identifiable information, including usernames, to protect data subjects.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? Yes
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? We include all relevant dataset details in the README of the data repository

A Additional Dataset Details

In the following, we briefly present additional details regarding the distribution of comments in the processed datasets. In Table 5, we present the distributions of comments that we collected per season, along with numbers of matches and players that were commented on. In Table 6, we present the 20 most viral players, along with the corresponding counts of comments, and in Figure 2, we present the overall distribution of comments per player. Further, in Table 7, we give the exact distribution of comments per club.

	Reddit			Twitter		
Season	**Posts**	**Players**	**Matches**	**Posts**	**Players**	**Matches**
17/18	46,658	542	294	177,288	633	478
18/19	54,358	534	280	150,563	649	467
19/20	39,203	529	222	144,364	614	455
20/21	63,998	616	271	175,047	692	464

Table 5: *Distribution of comments per season*. Next to the number of comments, we also include the number of player and matches that were commented on.

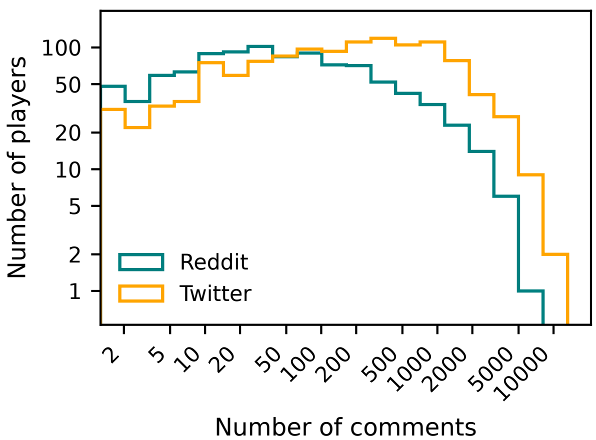

Figure 2: *Distribution of Comments Across Players.* We show histograms of comment counts per player, using 20 bins. We used a logarithmic scale, since we observed a long-tailed distribution of comments across players. Consequently, both axes are also scaled logarithmically.

Player Name	**Perceived Race**	**Reddit**	**Twitter**	**Total**
Harry Kane	White	3456	13,224	16,680
Mohamed Salah	Other	5418	10,562	15,980
Paul Pogba	Black	4115	7723	11,838
Raheem Sterling	Black	3357	7797	11,154
Sergio Agüero	Other	1911	7380	9291
Jamie Vardy	White	914	7987	8901
Marcus Rashford	Black	3470	5340	8810
Kevin De Bruyne	White	3096	5520	8616
Eden Hazard	White	1788	6768	8556
Sadio Mané	Black	3963	4347	8310
Heung-min Son	Other	2692	5252	7944
Wilfried Zaha	Black	757	6572	7329
David de Gea	White	2855	4018	6873
Romelu Lukaku	Black	2342	4079	6421
Olivier Giroud	White	1646	4756	6402
Willian	Black	1989	4115	6104
Riyad Mahrez	Other	1276	4716	5992
Álvaro Morata	Other	1946	3869	5815
Pierre-Emerick Aubameyang	Black	1161	4547	5708
Bruno Fernandes	Other	1884	3813	5697

Table 6: *Most viral players*. We show the number of comments per dataset for the 20 most viral players according to cumulated comments across both datasets.

Club Name	**Reddit**	**Twitter**	**Total**
Manchester United	34,248	58,620	92,868
Chelsea	25,021	64,061	89,082
Liverpool	31,141	55,483	86,624
Tottenham Hotspur	23,451	58,239	81,690
Manchester City	25,092	54,436	79,528
Arsenal	20,088	39,509	59,597
Everton	7374	33,486	40,860
Leicester City	4425	35,607	40,032
Southampton	2572	33,048	35,620
Crystal Palace	2900	31,836	34,736
Newcastle United	3323	29,218	32,541
Brighton & Hove Albion	3306	27,906	31,212
West Ham United	4004	25,764	29,768
Aston Villa	2347	18,893	21,240
Watford	1385	18,920	20,305
West Bromwich Albion	1092	12,247	13,339
Huddersfield Town	272	11,374	11,646
Fulham	1285	9926	11,211
Leeds United	1321	7326	8647
Wolverhampton Wanderers	3478	2975	6453
Stoke City	658	5406	6064
Burnley	2136	3733	5869
Bournemouth	1018	2798	3816
Swansea City	325	3264	3589
Sheffield United	1126	1360	2486
Cardiff City	299	1232	1531
Norwich City	530	630	1160

Table 7: *Distribution of comments per team.* We count a comment for a team, if the player that is commented on played for the corresponding team. Horizontal rule separates the *big six teams*, which are also the most viral, from the remaining 21 teams.

B LLM-based Word Sense Disambiguation for Soccer Supersenses

As discussed in Section 5, we use a combination of a dictionary and prompt-based LLMs to detect whether a post contains discussions of soccer supersenses. For 51 words (Table 8), we used LLaMa 3 8B to disambiguate whether it was being used to actually comment on a player's technical, physical, or mental attributes. We used the following prompt template for the LLaMa-based disambiguation to establish whether a social media post truly exhibited commentary on a player's technical, physical, or mental attributes.

```
In the following social media comment
about a soccer game, is the term [WORD] used
in the context of describing [CONTEXT] of
a soccer player? If yes, reply with '0',
otherwise, reply with '1'.
```

We replace `[WORD]` with the term that needs to be disambiguated, e.g., 'big' and `[CONTEXT]` with the corresponding attribute, e.g., "the physical attributes". The full list of terms we disambiguate is included in our online code supplement. We use default temperature settings (0.7) and access the model through huggingface.[11]

C Detailed Regression Results

Complementing Figure 1, we present the detailed regression results for Reddit in Table 9 and for Twitter in Table 10. In Table 15, we present results from regressions on additional toxicity-related variables. In Tables 13 and 14, we present regressions on supersenses, restricted to comments of positive and negative sentiment only.

In Table 11 and Table 12, we analyze the effect of using an indicator for big six clubs compared to encoding every single club in our main regressions. We observe that more club indicators generally improve model quality as measured by adjusted R^2, however, it does not have a notable effect on coefficient size and significance for the race-related attributes.

In addition to the 'toxicity' variable in the Perspective API, we also use the 'severe toxicity', 'identity attack', 'insult' and 'profanity' variables as dependent variables, since they can also be indicators of differential treatment based on race. For Twitter, as with toxicity (c.f. Section 7), results appear overall less strong, though results for insult and profanity still appear significant. For Reddit, all variables except for identity attack are also negatively associated with being White, i.e., Black players are more likely to receive severely toxic abuse and insults (Table 15).

D Matching Details

Finally, we present the list of matched players in Table 16.

Word	**Context**
awkward	technical ability
awkwardly	technical ability
big	height
bigger	height
biggest	height
defends	on-field actions
defended	on-field actions
defending	on-field actions
experience	experience level
experienced	experience level
fast	pace
faster	pace
fastest	pace
fight	mentality
fighter	mentality
fighting	mentality
fights	mentality
finished	shot conversion
finishes	shot conversion
finishing	shot conversion
fought	mentality
fragile	physical strength
hold	hold-up play
holding	hold-up play
inexperienced	experience level
power	physical strength
powerful	physical strength
quick	pace
quicker	pace
quickest	pace
quickly	pace
slow	pace
slower	pace
slowest	pace
small	height
smaller	height
smallest	height
speed	pace
speedy	pace
strong	physical strength
stronger	physical strength
strongest	physical strength
tiny	height
weak	physical strength
weaker	physical strength
weakest	physical strength
work	work rate
worked	work rate
save	goalkeeping actions
saves	goalkeeping actions
saved	goalkeeping actions

Table 8: *Ambiguous terms.* We show the full list of ambiguous terms ('word') that we used a prompt-based LLM to disambiguate and determine if it was actually used to describe players' soccer supersenses ('context').

[11] https://huggingface.co/meta-llama/Meta-Llama-3-8B

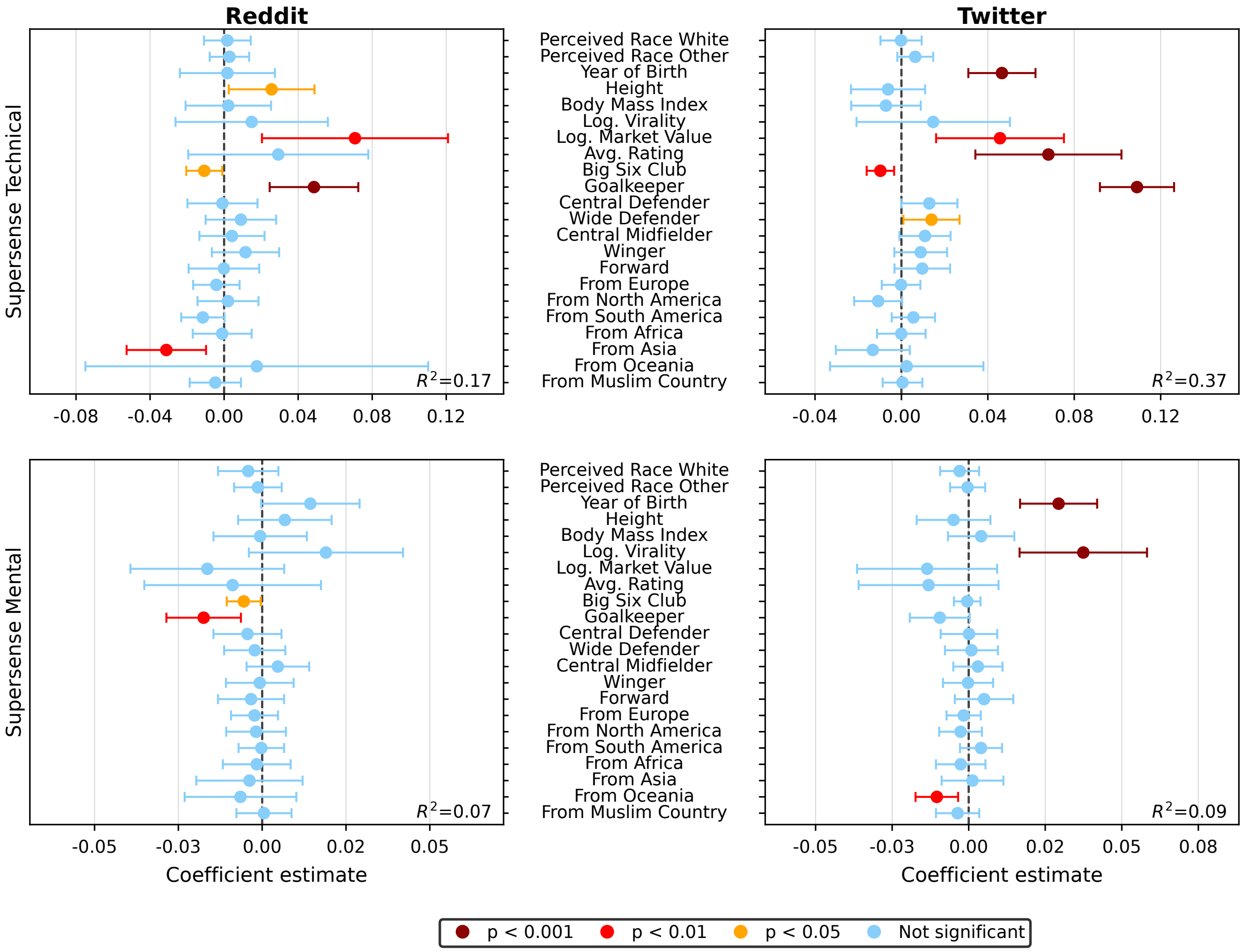

Figure 3: *Main regression results for technical and mental supersenses.* Each panel depicts coefficient estimates of independent variables, along with 95% confidence intervals, from the regressions on each combination of dependent variable (panel rows) and dataset (panel colums).

	Sentiment	Toxicity	Supersense Physical	Supersense Technical	Supersense Mental
Perceived Race Other	0.015	-0.012	-0.005	0.003	-0.001
	(0.015)	(0.008)	(0.003)	(0.005)	(0.004)
Perceived Race White	0.034	-0.028**	-0.007	0.002	-0.004
	(0.018)	(0.009)	(0.004)	(0.006)	(0.005)
Central Defender	-0.002	0.013	0.004	-0.001	-0.004
	(0.027)	(0.013)	(0.005)	(0.010)	(0.005)
Central Midfielder	0.023	0.001	0.002	0.004	0.005
	(0.025)	(0.011)	(0.004)	(0.009)	(0.005)
Forward	0.030	-0.008	0.002	-0.000	-0.003
	(0.029)	(0.012)	(0.005)	(0.010)	(0.005)
Goalkeeper	0.051	-0.004	-0.014**	0.049***	-0.017**
	(0.034)	(0.015)	(0.005)	(0.012)	(0.006)
Wide Defender	-0.016	0.010	0.000	0.009	-0.002
	(0.026)	(0.011)	(0.004)	(0.010)	(0.005)
Winger	0.005	0.006	0.005	0.012	-0.001
	(0.025)	(0.012)	(0.004)	(0.009)	(0.005)
Intercept	0.058	0.179***	0.015	0.010	0.038***
	(0.051)	(0.023)	(0.010)	(0.020)	(0.011)
Avg. Rating	0.080	-0.066	-0.014	0.029	-0.009
	(0.077)	(0.035)	(0.011)	(0.025)	(0.013)
Body Mass Index	0.033	-0.011	0.006	0.002	-0.001
	(0.031)	(0.016)	(0.005)	(0.012)	(0.007)
Year of Birth	0.100**	-0.042*	0.008	0.002	0.014
	(0.033)	(0.017)	(0.007)	(0.013)	(0.007)
Big Six Club	0.021	-0.025***	-0.003	-0.011*	-0.005*
	(0.013)	(0.006)	(0.003)	(0.005)	(0.003)
From Africa	0.039	-0.038***	-0.009	-0.001	-0.002
	(0.020)	(0.011)	(0.005)	(0.008)	(0.005)
From Asia	0.049*	-0.053***	-0.003	-0.031**	-0.004
	(0.023)	(0.015)	(0.005)	(0.011)	(0.008)
From Europe	0.006	-0.009	0.003	-0.004	-0.002
	(0.014)	(0.008)	(0.003)	(0.006)	(0.004)
From Muslim Country	-0.034	0.031*	0.006	-0.005	0.001
	(0.022)	(0.013)	(0.004)	(0.007)	(0.004)
From North America	-0.000	0.003	0.001	0.002	-0.002
	(0.021)	(0.011)	(0.005)	(0.008)	(0.005)
From Oceania	-0.094	-0.006	0.002	0.018	-0.006
	(0.118)	(0.019)	(0.003)	(0.047)	(0.008)
From South America	0.000	-0.013	-0.002	-0.011	-0.000
	(0.018)	(0.009)	(0.003)	(0.006)	(0.003)
Height	-0.064	0.032	0.018*	0.026*	0.007
	(0.034)	(0.017)	(0.007)	(0.012)	(0.007)
Log. Market Value	-0.058	0.021	-0.015	0.071**	-0.016
	(0.066)	(0.033)	(0.012)	(0.026)	(0.012)
Log. Virality	-0.151*	0.117***	0.016	0.015	0.019
	(0.065)	(0.030)	(0.010)	(0.021)	(0.012)
Observations	942	942	618	618	618
R^2	0.085	0.110	0.076	0.169	0.066
Adjusted R^2	0.064	0.088	0.042	0.138	0.031
Residual Std. Error	0.150 (df=919)	0.079 (df=919)	0.021 (df=595)	0.042 (df=595)	0.025 (df=595)
F Statistic	2.619*** (df=22; 919)	4.948*** (df=22; 919)	3.262*** (df=22; 595)	5.413*** (df=22; 595)	3.273*** (df=22; 595)
Note:					*p<0.05; **p<0.01; ***p<0.001

Table 9: *Detailed regression results for Reddit data.* Complementing the illustration in Figures 1 and 3, we present the exact numerical values of the coefficients and their deviations.

	Sentiment	Toxicity	Supersense Physical	Supersense Technical	Supersense Mental
Perceived Race Other	0.018	-0.005	0.002	0.006	-0.000
	(0.013)	(0.006)	(0.002)	(0.004)	(0.003)
Perceived Race White	0.019	-0.014*	0.001	-0.000	-0.003
	(0.014)	(0.006)	(0.002)	(0.005)	(0.003)
Central Defender	-0.013	0.018***	0.006*	0.013	0.000
	(0.017)	(0.005)	(0.003)	(0.007)	(0.005)
Central Midfielder	-0.010	0.018**	0.003	0.011	0.003
	(0.018)	(0.006)	(0.002)	(0.006)	(0.004)
Forward	0.001	0.006	0.000	0.010	0.005
	(0.017)	(0.005)	(0.003)	(0.007)	(0.005)
Goalkeeper	0.025	0.016*	-0.001	0.109***	-0.009
	(0.021)	(0.006)	(0.003)	(0.009)	(0.005)
Wide Defender	-0.017	0.021***	0.004	0.014*	0.001
	(0.016)	(0.004)	(0.003)	(0.007)	(0.004)
Winger	-0.005	0.008	0.005	0.009	-0.000
	(0.015)	(0.004)	(0.003)	(0.006)	(0.004)
Intercept	0.139***	0.062***	0.008	-0.014	0.038***
	(0.035)	(0.011)	(0.006)	(0.014)	(0.009)
Avg. Rating	0.091	-0.030	-0.016**	0.068***	-0.013
	(0.052)	(0.019)	(0.006)	(0.017)	(0.012)
Body Mass Index	-0.004	0.004	-0.002	-0.007	0.004
	(0.024)	(0.007)	(0.003)	(0.008)	(0.006)
Year of Birth	0.198***	-0.009	0.019***	0.047***	0.029***
	(0.025)	(0.010)	(0.003)	(0.008)	(0.006)
Big Six Club	0.049***	-0.012***	-0.000	-0.010**	-0.000
	(0.008)	(0.003)	(0.001)	(0.003)	(0.002)
From Africa	-0.011	-0.010	0.006*	-0.000	-0.003
	(0.016)	(0.006)	(0.003)	(0.006)	(0.004)
From Asia	0.002	-0.013	0.000	-0.013	0.001
	(0.022)	(0.007)	(0.003)	(0.009)	(0.005)
From Europe	-0.023	0.002	-0.001	-0.000	-0.002
	(0.014)	(0.005)	(0.002)	(0.005)	(0.003)
From Muslim Country	0.015	0.001	-0.003	0.000	-0.004
	(0.014)	(0.004)	(0.002)	(0.005)	(0.004)
From North America	0.004	-0.013*	0.001	-0.011	-0.003
	(0.015)	(0.005)	(0.003)	(0.006)	(0.004)
From Oceania	0.031	-0.033**	-0.002	0.002	-0.010**
	(0.040)	(0.010)	(0.005)	(0.018)	(0.004)
From South America	-0.019	0.002	0.001	0.006	0.004
	(0.018)	(0.009)	(0.002)	(0.005)	(0.004)
Height	-0.033	0.007	0.002	-0.006	-0.005
	(0.025)	(0.009)	(0.004)	(0.009)	(0.006)
Log. Market Value	-0.237***	-0.005	-0.005	0.046**	-0.014
	(0.043)	(0.014)	(0.006)	(0.015)	(0.012)
Log. Virality	0.086	0.056***	0.014*	0.015	0.037***
	(0.048)	(0.017)	(0.006)	(0.018)	(0.011)
Observations	1007	1007	891	891	891
R^2	0.163	0.073	0.099	0.371	0.094
Adjusted R^2	0.144	0.052	0.076	0.355	0.071
Residual Std. Error	0.117 (df=984)	0.043 (df=984)	0.015 (df=868)	0.039 (df=868)	0.026 (df=868)
F Statistic	7.355*** (df=22; 984)	6.137*** (df=22; 984)	3.873*** (df=22; 868)	20.561*** (df=22; 868)	5.344*** (df=22; 868)
Note:					*p<0.05; **p<0.01; ***p<0.001

Table 10: *Detailed regression results for Twitter data.* Complementing the illustration in Figures 1 and 3, we present the exact numerical values of the coefficients and their deviations.

	Big Six Club Only						All Clubs					
	β_{white}	p_{white}	β_{other}	p_{other}	R^2	$\bar{R}^2$	β_{white}	p_{white}	β_{other}	p_{other}	R^2	$\bar{R}^2$
Sentiment Score	0.03	0.06	0.02	0.32	0.09	0.06	0.03	0.13	0.01	0.35	0.11	0.07
Toxicity Score	-0.03	0.00	-0.01	0.14	0.11	0.09	-0.03	0.00	-0.01	0.17	0.16	0.11
Supersense Physical	-0.01	0.12	-0.00	0.15	0.08	0.04	-0.01	0.15	-0.01	0.17	0.12	0.04
Supersense Technical	0.00	0.79	0.00	0.60	0.17	0.14	0.00	0.82	0.00	0.68	0.23	0.17
Supersense Mental	-0.00	0.36	-0.00	0.73	0.07	0.03	-0.01	0.26	-0.00	0.68	0.12	0.05

Table 11: *Comparison of club encodings on Reddit data.* We compare results of the regression model which uses a big six club-indicator variable against a regression model using an indicator variable for every club. β_{white} and β_{other} indicate regression coefficient for the perceived race attributes, p_{white} and p_{other} the corresponding p-values. $\bar{R}^2$ denotes the adjusted R^2 value. We observe that introducing indicators for every single club slightly increase the adjusted R^2 scores, but only hardly impact value or significance of the perceived race-related regression coefficients.

	Big Six Club Only						All Clubs					
	β_{white}	p_{white}	β_{other}	p_{other}	R^2	$\bar{R}^2$	β_{white}	p_{white}	β_{other}	p_{other}	R^2	$\bar{R}^2$
Sentiment Score	0.02	0.18	0.02	0.16	0.16	0.14	0.01	0.36	0.01	0.31	0.22	0.18
Toxicity Score	-0.01	0.02	-0.00	0.39	0.07	0.05	-0.01	0.04	-0.00	0.38	0.17	0.13
Supersense Physical	0.00	0.51	0.00	0.46	0.10	0.08	0.00	0.69	0.00	0.70	0.15	0.10
Supersense Technical	-0.00	0.97	0.01	0.13	0.37	0.36	-0.00	0.64	0.01	0.17	0.42	0.39
Supersense Mental	-0.00	0.36	-0.00	0.90	0.09	0.07	-0.00	0.32	-0.00	0.67	0.16	0.11

Table 12: *Comparison of club encodings on Twitter data.* We compare results of the regression model which uses a big six club-indicator variable against a regression model using an indicator variable for every club. β_{white} and β_{other} indicate regression coefficient for the perceived race attributes, p_{white} and p_{other} the corresponding p-values. $\bar{R}^2$ denotes the adjusted R^2 value. We observe that introducing indicators for every single club slightly increase the adjusted R^2 scores, but only hardly impact value or significance of the perceived race-related regression coefficients.

	Positive Sentiment			Negative Sentiment		
	Supersense Physical	**Supersense Technical**	**Supersense Mental**	**Supersense Physical**	**Supersense Technical**	**Supersense Mental**
Perceived Race Other	-0.009	-0.006	0.004	0.007	0.013	-0.011
	(0.006)	(0.011)	(0.008)	(0.007)	(0.011)	(0.011)
Perceived Race White	-0.007	0.013	0.004	0.003	0.004	-0.019
	(0.006)	(0.013)	(0.008)	(0.007)	(0.013)	(0.015)
Central Defender	-0.005	0.010	-0.001	0.014	-0.008	-0.003
	(0.010)	(0.019)	(0.012)	(0.008)	(0.021)	(0.015)
Central Midfielder	-0.001	0.006	0.008	-0.002	0.020	0.020
	(0.010)	(0.017)	(0.009)	(0.005)	(0.021)	(0.016)
Forward	-0.002	0.011	0.015	0.013	0.003	-0.001
	(0.011)	(0.018)	(0.011)	(0.008)	(0.022)	(0.016)
Goalkeeper	-0.020	0.164^{***}	-0.011	-0.005	-0.013	-0.030
	(0.011)	(0.024)	(0.012)	(0.009)	(0.024)	(0.016)
Wide Defender	-0.008	0.018	0.001	0.004	0.003	-0.007
	(0.010)	(0.017)	(0.010)	(0.006)	(0.022)	(0.015)
Winger	-0.000	0.016	-0.001	0.008	0.028	-0.003
	(0.010)	(0.018)	(0.009)	(0.006)	(0.022)	(0.015)
Intercept	0.036	0.021	-0.001	-0.008	0.008	0.139^{***}
	(0.020)	(0.041)	(0.020)	(0.015)	(0.043)	(0.039)
Avg. Rating	-0.030	0.034	0.025	-0.010	0.047	-0.002
	(0.019)	(0.053)	(0.035)	(0.025)	(0.047)	(0.034)
Body Mass Index	-0.005	-0.019	-0.003	0.031^{*}	0.045	-0.012
	(0.010)	(0.024)	(0.014)	(0.013)	(0.024)	(0.020)
Year of Birth	0.008	0.023	0.016	0.020	0.010	-0.005
	(0.012)	(0.028)	(0.016)	(0.018)	(0.037)	(0.024)
Big Six Club	-0.004	-0.011	0.005	-0.000	-0.011	-0.011
	(0.005)	(0.011)	(0.007)	(0.005)	(0.009)	(0.007)
From Africa	-0.012	-0.008	0.007	0.008	0.005	-0.029
	(0.006)	(0.016)	(0.010)	(0.009)	(0.016)	(0.017)
From Asia	-0.017^{*}	-0.036	0.022	0.029	-0.045^{*}	-0.044^{*}
	(0.008)	(0.022)	(0.019)	(0.019)	(0.019)	(0.018)
From Europe	-0.001	-0.003	0.008	0.001	-0.011	-0.017
	(0.006)	(0.013)	(0.008)	(0.008)	(0.013)	(0.013)
From Muslim Country	0.014^{*}	0.016	0.011	0.002	-0.024	0.011
	(0.006)	(0.014)	(0.011)	(0.010)	(0.015)	(0.013)
From North America	0.002	0.005	0.006	-0.004	0.017	-0.011
	(0.007)	(0.017)	(0.008)	(0.006)	(0.020)	(0.016)
From Oceania	-0.001	0.044	-0.001	-0.005	0.039	-0.022
	(0.013)	(0.089)	(0.016)	(0.008)	(0.043)	(0.029)
From South America	0.005	-0.014	0.004	0.004	-0.016	-0.012
	(0.005)	(0.013)	(0.007)	(0.007)	(0.012)	(0.013)
Height	0.015	0.011	0.009	0.011	0.031	-0.019
	(0.011)	(0.026)	(0.017)	(0.015)	(0.024)	(0.022)
Log. Market Value	-0.014	0.110^{*}	0.023	-0.024	0.054	-0.046
	(0.020)	(0.055)	(0.026)	(0.034)	(0.068)	(0.040)
Log. Virality	0.028	-0.008	-0.020	0.014	-0.014	0.011
	(0.019)	(0.051)	(0.035)	(0.023)	(0.046)	(0.031)
Observations	618	618	618	618	618	618
R^2	0.047	0.216	0.047	0.061	0.049	0.054
Adjusted R^2	0.011	0.187	0.012	0.027	0.014	0.019
Residual Std. Error	0.037 (df=595)	0.089 (df=595)	0.053 (df=595)	0.044 (df=595)	0.092 (df=595)	0.077 (df=595)
F Statistic	1.763^{*} (df=22; 595)	6.822^{***} (df=22; 595)	2.425^{***} (df=22; 595)	2.272^{***} (df=22; 595)	2.229^{**} (df=22; 595)	1.898^{**} (df=22; 595)
Note:						*p<0.05; **p<0.01; ***p<0.001

Table 13: *Results of regressions on supersenses on Reddit data, restricted to positive and negative sentiment comments.*

	Positive Sentiment			Negative Sentiment		
	Supersense Physical	**Supersense Technical**	**Supersense Mental**	**Supersense Physical**	**Supersense Technical**	**Supersense Mental**
Perceived Race Other	-0.002	0.013*	0.000	0.011	0.001	0.002
	(0.003)	(0.007)	(0.005)	(0.007)	(0.009)	(0.008)
Perceived Race White	-0.001	0.002	0.000	0.009	-0.001	0.001
	(0.004)	(0.007)	(0.006)	(0.006)	(0.011)	(0.009)
Central Defender	0.009*	0.002	-0.005	0.017*	0.036**	0.002
	(0.005)	(0.011)	(0.008)	(0.007)	(0.011)	(0.011)
Central Midfielder	0.004	0.005	0.001	0.010	0.034***	0.004
	(0.004)	(0.010)	(0.007)	(0.006)	(0.009)	(0.010)
Forward	0.003	0.004	0.001	0.014	0.042***	0.006
	(0.004)	(0.010)	(0.008)	(0.009)	(0.011)	(0.012)
Goalkeeper	0.002	0.177***	-0.015	0.008	0.090***	-0.021
	(0.005)	(0.015)	(0.009)	(0.008)	(0.014)	(0.012)
Wide Defender	0.007	0.004	-0.004	0.013	0.046***	0.008
	(0.004)	(0.010)	(0.008)	(0.007)	(0.011)	(0.010)
Winger	0.007	0.002	0.001	0.009	0.049***	0.006
	(0.004)	(0.010)	(0.007)	(0.008)	(0.013)	(0.013)
Intercept	0.010	-0.015	0.060***	0.013	-0.004	0.063*
	(0.010)	(0.021)	(0.018)	(0.017)	(0.026)	(0.028)
Avg. Rating	-0.042***	0.080**	-0.041	0.021	0.043	-0.065*
	(0.012)	(0.028)	(0.025)	(0.045)	(0.030)	(0.025)
Body Mass Index	0.000	-0.005	0.002	-0.008	-0.023	-0.007
	(0.006)	(0.014)	(0.010)	(0.009)	(0.016)	(0.016)
Year of Birth	0.026***	0.063***	0.024*	0.011	0.022	-0.012
	(0.006)	(0.012)	(0.011)	(0.012)	(0.016)	(0.014)
Big Six Club	-0.001	-0.015**	-0.004	-0.002	-0.011	0.003
	(0.002)	(0.005)	(0.003)	(0.003)	(0.009)	(0.008)
From Africa	0.004	0.001	-0.001	0.019*	-0.020	0.007
	(0.004)	(0.009)	(0.007)	(0.009)	(0.011)	(0.011)
From Asia	0.006	-0.016	0.006	0.004	-0.022	-0.004
	(0.006)	(0.013)	(0.009)	(0.010)	(0.015)	(0.012)
From Europe	0.000	0.004	0.004	-0.002	-0.010	-0.008
	(0.003)	(0.007)	(0.005)	(0.005)	(0.009)	(0.009)
From Muslim Country	0.002	-0.003	-0.001	-0.018**	0.004	-0.015
	(0.004)	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)
From North America	0.005	-0.007	-0.007	0.006	-0.005	-0.000
	(0.005)	(0.009)	(0.007)	(0.010)	(0.013)	(0.009)
From Oceania	-0.015*	0.004	-0.018**	-0.003	-0.021	-0.008
	(0.006)	(0.020)	(0.007)	(0.008)	(0.011)	(0.011)
From South America	0.003	0.005	0.007	-0.001	-0.011	0.009
	(0.003)	(0.008)	(0.006)	(0.005)	(0.009)	(0.007)
Height	0.002	-0.019	-0.004	-0.006	-0.001	0.017
	(0.007)	(0.015)	(0.011)	(0.015)	(0.018)	(0.014)
Log. Market Value	0.008	0.087***	0.034	-0.012	0.095***	0.034
	(0.009)	(0.025)	(0.018)	(0.031)	(0.027)	(0.024)
Log. Virality	0.017	-0.005	0.009	-0.012	-0.013	0.024
	(0.011)	(0.030)	(0.023)	(0.021)	(0.040)	(0.039)
Observations	891	891	891	886	886	886
R^2	0.084	0.396	0.048	0.026	0.086	0.030
Adjusted R^2	0.061	0.380	0.024	0.001	0.063	0.006
Residual Std. Error	0.026 (df=868)	0.061 (df=868)	0.048 (df=868)	0.047 (df=863)	0.079 (df=863)	0.069 (df=863)
F Statistic	4.004*** (df=22; 868)	17.545*** (df=22; 868)	2.587*** (df=22; 868)	0.910 (df=22; 863)	6.171*** (df=22; 863)	2.709*** (df=22; 863)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 14: *Results of regressions on supersenses on Twitter data, restricted to positive and negative sentiment comments.*

	Reddit				Twitter			
	Severe Toxicity	**Identity Attack**	**Insult**	**Profanity**	**Severe Toxicity**	**Identity Attack**	**Insult**	**Profanity**
Perceived Race Other	-0.004	-0.000	-0.008	-0.014	-0.001	0.001	-0.004	-0.004
	(0.003)	(0.002)	(0.006)	(0.008)	(0.001)	(0.001)	(0.005)	(0.004)
Perceived Race White	-0.008**	-0.003	-0.019**	-0.029***	-0.003	-0.000	-0.009*	-0.011*
	(0.003)	(0.002)	(0.007)	(0.009)	(0.001)	(0.001)	(0.005)	(0.004)
Central Defender	0.006	0.002	0.011	0.011	0.004**	0.001	0.013***	0.015***
	(0.004)	(0.003)	(0.009)	(0.014)	(0.001)	(0.001)	(0.004)	(0.004)
Central Midfielder	0.001	-0.000	0.001	-0.002	0.004*	0.001	0.013**	0.014**
	(0.003)	(0.002)	(0.007)	(0.011)	(0.002)	(0.001)	(0.005)	(0.005)
Forward	-0.002	-0.001	-0.009	-0.004	0.001	0.000	0.004	0.005
	(0.004)	(0.003)	(0.008)	(0.013)	(0.001)	(0.001)	(0.003)	(0.004)
Goalkeeper	-0.001	0.001	-0.007	0.004	0.003	0.000	0.010*	0.013**
	(0.004)	(0.003)	(0.010)	(0.016)	(0.002)	(0.001)	(0.004)	(0.005)
Wide Defender	0.001	0.003	0.012	0.000	0.004***	0.002*	0.015***	0.017***
	(0.004)	(0.003)	(0.007)	(0.012)	(0.001)	(0.001)	(0.003)	(0.003)
Winger	0.002	0.001	0.002	0.001	0.002	0.000	0.006*	0.007*
	(0.004)	(0.002)	(0.008)	(0.013)	(0.001)	(0.001)	(0.003)	(0.003)
Intercept	0.029***	0.030***	0.116***	0.133***	0.008***	0.010***	0.036***	0.040***
	(0.007)	(0.006)	(0.016)	(0.023)	(0.003)	(0.002)	(0.008)	(0.008)
Avg. Rating	-0.019	-0.010	-0.064**	-0.042	-0.003	-0.001	-0.021	-0.022
	(0.011)	(0.008)	(0.023)	(0.035)	(0.007)	(0.005)	(0.013)	(0.014)
Body Mass Index	-0.003	-0.006	-0.007	-0.014	0.001	0.001	0.002	0.002
	(0.005)	(0.004)	(0.012)	(0.014)	(0.002)	(0.002)	(0.005)	(0.005)
Year of Birth	-0.011*	-0.008*	-0.034**	-0.035*	-0.001	-0.001	-0.004	-0.006
	(0.005)	(0.003)	(0.012)	(0.017)	(0.003)	(0.002)	(0.008)	(0.007)
Big Six Club	-0.007***	-0.003*	-0.021***	-0.018**	-0.003***	-0.001	-0.008***	-0.009***
	(0.002)	(0.001)	(0.004)	(0.006)	(0.001)	(0.001)	(0.002)	(0.002)
From Africa	-0.010**	-0.004*	-0.030***	-0.031**	-0.002	-0.000	-0.008	-0.008
	(0.003)	(0.002)	(0.008)	(0.011)	(0.002)	(0.001)	(0.005)	(0.004)
From Asia	-0.011*	-0.006	-0.041***	-0.036**	-0.002	-0.002	-0.009*	-0.006
	(0.004)	(0.003)	(0.011)	(0.014)	(0.002)	(0.001)	(0.005)	(0.005)
From Europe	0.001	-0.004	-0.008	-0.003	0.001	-0.002*	0.002	0.003
	(0.003)	(0.002)	(0.006)	(0.007)	(0.001)	(0.001)	(0.004)	(0.003)
From Muslim Country	0.009	0.006***	0.023**	0.027*	0.000	0.001	0.001	0.002
	(0.005)	(0.002)	(0.008)	(0.013)	(0.001)	(0.001)	(0.003)	(0.003)
From North America	0.002	0.001	-0.001	0.008	-0.003	-0.001*	-0.009*	-0.011**
	(0.004)	(0.002)	(0.008)	(0.011)	(0.002)	(0.001)	(0.004)	(0.004)
From Oceania	-0.002	-0.002	-0.015	-0.004	-0.007**	-0.003**	-0.022**	-0.025***
	(0.006)	(0.002)	(0.013)	(0.025)	(0.002)	(0.001)	(0.008)	(0.007)
From South America	-0.001	-0.001	-0.008	-0.006	0.000	-0.001	0.004	0.004
	(0.002)	(0.002)	(0.006)	(0.008)	(0.002)	(0.001)	(0.008)	(0.007)
Height	0.009	-0.000	0.024	0.029	0.001	-0.001	0.007	0.004
	(0.006)	(0.003)	(0.012)	(0.017)	(0.002)	(0.001)	(0.007)	(0.007)
Log. Market Value	0.007	-0.001	0.026	0.025	-0.003	-0.005	-0.006	0.000
	(0.012)	(0.005)	(0.022)	(0.033)	(0.004)	(0.003)	(0.010)	(0.010)
Log. Virality	0.027*	0.012**	0.080***	0.081**	0.010*	0.006**	0.039**	0.046***
	(0.012)	(0.004)	(0.020)	(0.030)	(0.005)	(0.002)	(0.013)	(0.012)
Observations	942	942	942	942	1007	1007	1007	1007
R^2	0.068	0.057	0.118	0.081	0.042	0.041	0.060	0.085
Adjusted R^2	0.046	0.034	0.097	0.060	0.021	0.020	0.039	0.064
Residual Std. Error	0.026 (df=919)	0.015 (df=919)	0.056 (df=919)	0.078 (df=919)	0.012 (df=984)	0.006 (df=984)	0.033 (df=984)	0.032 (df=984)
F Statistic	6.230*** (df=22; 919)	2.678*** (df=22; 919)	6.383*** (df=22; 919)	4.646*** (df=22; 919)	6.197*** (df=22; 984)	3.077*** (df=22; 984)	6.027*** (df=22; 984)	7.986*** (df=22; 984)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 15: *Regression results for additional toxicity-related variables.*

Black Player	White Player	Mahalanobis Distance
Jermain Defoe	Billy Sharp	1.848035
Patrice Evra	Leighton Baines	2.147109
Vincent Kompany	Phil Jones	1.661798
Wes Morgan	Phil Jagielka	1.187269
Yaya Touré	Charlie Adam	2.533960
Ashley Young	Alberto Moreno	2.582167
Aaron Lennon	Robbie Brady	2.155753
Ryan Babel	Stuart Dallas	1.869811
Mousa Dembélé	Aaron Ramsey	1.374583
Fernandinho	Ander Herrera	1.970276
Michel Vorm	Willy Caballero	2.254699
Antonio Valencia	Nacho Monreal	1.546437
David McGoldrick	Billy Sharp	2.428342
Gaëtan Bong	Martin Kelly	1.570642
Ryan Bertrand	Javier Manquillo	2.661236
André Ayew	Robbie Brady	1.527986
Moussa Sissoko	Jordan Henderson	1.996437
Bakary Sako	James McClean	1.533318
Daniel Sturridge	Ashley Barnes	2.469341
Mamadou Sakho	Lewis Dunk	1.316156
Jason Puncheon	James McCarthy	1.944056
Angelo Ogbonna	Federico Fernández	0.942504
Georginio Wijnaldum	Ander Herrera	1.309627
Danny Rose	Kieran Trippier	2.004873
Christian Benteke	Chris Wood	1.498475
Fabian Delph	Jack Wilshere	1.793995
Patrick van Aanholt	Séamus Coleman	1.060281
Willian	Matt Ritchie	2.355247
Pierre-Emerick Aubameyang	Gonzalo Higuaín	2.084693
Denis Odoi	Javier Manquillo	1.416777
Victor Moses	Matt Ritchie	2.394443
Mame Diouf	Sam Vokes	1.142358
Odion Ighalo	Lucas Pérez	2.333534
Étienne Capoue	Chris Brunt	1.732491
Troy Deeney	Jamie Vardy	1.582560
Cheikhou Kouyaté	Morgan Schneiderlin	1.484281
Danny Welbeck	Danny Ings	2.225557
Ashley Williams	Scott Dann	1.482592
Mohamed Diamé	Dale Stephens	1.195758
Francis Coquelin	Danny Drinkwater	2.027044
Yannick Bolasie	Andriy Yarmolenko	2.113864
Matt Phillips	André Schürrle	1.985699
Victor Wanyama	Danny Drinkwater	1.391828
Salomón Rondón	Patrick Bamford	1.207553
Christian Kabasele	Ciaran Clark	0.933856
Wilfried Bony	Teemu Pukki	1.739114
Joel Matip	Aymeric Laporte	1.214318
Nathaniel Clyne	Konstantinos Tsimikas	2.009335
Antonio Rüdiger	Eric Dier	0.611216
Eliaquim Mangala	Toby Alderweireld	1.726116
Joshua King	Ashley Barnes	1.767107
Cuco Martina	Chris Basham	0.779249
Alexandre Lacazette	Danny Ings	1.587487
Romelu Lukaku	Harry Kane	3.028280
Pedro Obiang	Davy Pröpper	1.089342
Michail Antonio	Shane Long	2.343412
Jordan Ayew	Solly March	0.877623
Nampalys Mendy	Aaron Mooy	1.746664
Allan Nyom	Christian Fuchs	1.163665
Andre Gray	Shane Long	2.077134
Alphonse Areola	Jack Butland	0.453296
Paul Pogba	Scott McTominay	2.007000
Callum Wilson	Jay Rodríguez	1.441892
Idrissa Gueye	Yohan Cabaye	1.240751
Serge Aurier	Luke Shaw	1.813730
Abdoulaye Doucouré	Stuart Armstrong	1.219833
Saido Berahino	Florin Andone	1.795981
Nathan Redmond	Matt Ritchie	1.959190
Bertrand Traoré	Jarrod Bowen	1.186026
Raheem Sterling	Eden Hazard	1.962379
Virgil van Dijk	Harry Maguire	1.038842
Ryan Fredericks	Charlie Taylor	0.442732
Willy Boly	Shane Duffy	1.238433
Wilfried Zaha	Harvey Barnes	1.965420
Jetro Willems	George Baldock	1.625137
José Izquierdo	Matt Ritchie	1.826060
Divock Origi	Kai Havertz	2.494522
Ivan Cavaleiro	Matt Ritchie	1.851308
Terence Kongolo	Adam Webster	1.298142
Benjamin Mendy	Kieran Tierney	1.242653
Jeffrey Schlupp	Marc Albrighton	1.216829
Kurt Zouma	Shkodran Mustafi	2.355730
Cyrus Christie	Joël Veltman	1.128917
Djibril Sidibé	Luke Ayling	1.194544
Jürgen Locadia	Charlie Austin	1.796463
Mario Lemina	Oriol Romeu	0.620945
Nathan Aké	Ben Davies	0.894639
Jean Michaël Seri	Adrien Silva	1.367325
Michy Batshuayi	Kai Havertz	2.176361
Sébastien Haller	Marko Arnautovic	1.716215
Arthur Masuaku	Matt Targett	1.359172

Black Player	White Player	Mahalanobis Distance
Tiemoué Bakayoko	Christian Eriksen	1.601531
Anthony Martial	Kai Havertz	2.025478
Jamaal Lascelles	Michael Keane	0.908210
Bobby De Cordova-Reid	Ryan Fraser	1.686025
Christian Atsu	Matt Ritchie	1.697446
Fred	Jack Wilshere	1.352095
Leroy Sané	Christian Pulisic	1.999351
Jordon Ibe	Stuart Dallas	2.923713
Dwight Gayle	Shane Long	1.672544
Jacob Murphy	Anthony Gordon	1.914926
Sadio Mané	Christian Pulisic	2.141729
Ruben Loftus-Cheek	Pierre-Emile Højbjerg	1.034056
Adama Traoré	Jack Grealish	1.470191
Oumar Niasse	Ashley Barnes	0.847256
Daniel Amartey	Jack Stephens	1.098105
Ricardo Pereira	Lucas Digne	1.309725
Zack Steffen	Dean Henderson	1.588657
N'Golo Kanté	Mateo Kovacic	2.034094
Fabinho	Ross Barkley	1.185143
Matheus Pereira	James Maddison	1.816494
Thomas Partey	Mateo Kovacic	1.634138
Alex Iwobi	Timo Werner	1.130835
Renato Sanches	Alexis Mac Allister	1.301271
Tosin Adarabioyo	Joe Rodon	1.000844
Marcus Rashford	Christian Pulisic	1.390447
Jefferson Lerma	Oliver Norwood	1.776646
Kevin Danso	Alfie Mawson	2.782265
Mbaye Diagne	Oli McBurnie	1.614474
Issa Diop	Joachim Andersen	1.368078
Allan Saint-Maximin	Harvey Barnes	2.558184
Wilfred Ndidi	Kalvin Phillips	1.227032
Steven Bergwijn	Harry Wilson	1.195978
Ainsley Maitland-Niles	Brandon Williams	1.120351
Eric Bailly	Andreas Christensen	0.839213
Yerry Mina	Jan Bednarek	1.684782
Aboubakar Kamara	Aaron Connolly	1.384484
Lys Mousset	Fábio Silva	1.656687
Demarai Gray	Oliver Burke	1.685931
Kelechi Iheanacho	Florin Andone	2.011856
Isaac Success	Florin Andone	1.956157
Naby Keïta	Jordan Henderson	1.884031
Fikayo Tomori	Rob Holding	1.886521
Percy Tau	Anthony Gordon	2.154282
Trent Alexander-Arnold	Luke Shaw	1.390253
Timothy Fosu-Mensah	Neco Williams	1.181989
Philip Billing	Tomas Soucek	1.643479
Ollie Watkins	Luciano Vietto	3.163704
Tammy Abraham	Kai Havertz	1.244938
Joelinton	Tom Davies	1.860896
Gedson Fernandes	Conor Gallagher	1.174048
Nathan Tella	Anthony Gordon	2.424794
Eddie Nketiah	Florin Andone	2.325030
Reiss Nelson	Patrick Roberts	2.485232
Joe Willock	Curtis Jones	1.004826
Kyle Walker-Peters	Oleksandr Zinchenko	0.874994
Davinson Sánchez	Victor Lindelöf	0.872286
Axel Tuanzebe	Konstantinos Tsimikas	2.068688
Nicolas Pépé	Harvey Barnes	2.357566
Japhet Tanganga	Calum Chambers	1.743364
Mason Holgate	Jonny Evans	2.687619
Frank Anguissa	Leander Dendoncker	1.607324
Gabriel Jesus	Harry Kane	2.521982
Moise Kean	Florin Andone	2.534625
Bernardo	Matty Cash	1.057849
Wesley Moraes	Alexander Sørloth	1.937604
Callum Hudson-Odoi	Phil Foden	1.629895
Ryan Sessegnon	Konstantinos Tsimikas	2.095166
Robert Sánchez	Aaron Ramsdale	1.684252
Yan Valery	Emil Krafth	2.057780
Ademola Lookman	Neal Maupay	2.068920
Rhian Brewster	Troy Parrott	2.023748
Ismaïla Sarr	Harvey Barnes	1.312827
Yves Bissouma	Pascal Groß	1.376894
Keinan Davis	Andy Carroll	3.237661
James Justin	Jonjoe Kenny	1.711309
Ezri Konsa	Alfie Mawson	1.101324
Jean-Philippe Mateta	Alexander Sørloth	1.603489
Bukayo Saka	Anthony Gordon	3.032195
Tino Anjorin	Emile Smith Rowe	2.457512
Edouard Mendy	Bernd Leno	0.936957
Michael Obafemi	Neal Maupay	2.588272
Tanguy Ndombélé	Jordan Henderson	1.763251
Reece James	Ben Chilwell	1.442731
Wesley Fofana	Wesley Hoedt	1.724939
Aaron Wan-Bissaka	Andrew Robertson	2.308015
Eberechi Eze	James Maddison	1.849443
Moussa Djenepo	Anthony Gordon	1.989671
Jeremy Ngakia	Phil Bardsley	5.472293
Tariq Lamptey	Alberto Moreno	2.333665
Amad Diallo	Anthony Gordon	3.407083
Mohammed Salisu	Fabian Schär	2.209129

Table 16: **List of Matched Players**. Third column indicates Mahalanobis distance with respect to the attributes that we matched on. Rows colored in red indicate matches which were discarded because Mahalanobis distance was too large, i.e., larger than 1.75. In total, we obtained 100 matches.