

Digital Social Pressure, Managerial Decision-Making and Worker Performance: Evidence from Professional Football

by **Andy Chung, Tin Cheuk Leung and J. James Reade**

Discussion Paper No. 2026-10

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

Digital Social Pressure, Managerial Decision-Making and Worker Performance: Evidence from Professional Football

Andy Chung
University of Reading
Global Labor Organization

Tin Cheuk Leung
Wake Forest University

James Reade
University of Reading

August 2026

Abstract

Digital platforms expose workers and managers to immediate, public, and persistent evaluations from external audiences, creating a form of social pressure that may shape workplace decisions. We examine whether such digital social pressure predicts subsequent opportunity allocation and worker performance, using professional football as a high-frequency empirical setting. We link player-match data from Europe's five major leagues to post-match Twitter commentary and construct measures of negative sentiment, toxicity, and racial abuse. Comparing players with themselves over time and teammates in the same current match, we find that greater negative sentiment is associated with a lower probability of appearing in the team's next match. Toxicity shows a similar but less stable association, while racial-abuse intensity is imprecisely estimated when the measures enter jointly. By contrast, associations with performance conditional on a subsequent appearance are small and imprecise. Overall, the findings suggest that digital social pressure is reflected more strongly in the allocation of subsequent opportunities than in worker performance, showing how public evaluations generated outside formal organizational boundaries can become part of the reputational environment surrounding internal managerial decisions.

Keywords: digital social pressure, online criticism, managerial decision-making, worker performance, social media, professional football, racial abuse

JEL codes: J24, M54, D91, Z22, J71

1 Introduction

Public evaluations can be associated with workplace outcomes even when the evaluators have no formal authority. They may alter workers' incentives and managers' willingness to allocate future opportunities. Economic research shows that social approval and reputational concerns shape conformity (Bernheim, 1994), peer monitoring is related to workplace effort and productivity (Mas and Moretti, 2009; Bandiera et al., 2010; Cornelissen et al., 2017), and social pressure changes charitable giving and professional decisions (DellaVigna et al., 2012; Garicano et al., 2005). This literature generally examines pressure from peers, supervisors, clients, or face-to-face audiences. Less is known about whether public evaluations produced outside an organization predict the outcomes of the individual workers being evaluated.

Online platforms make this question especially salient. They distribute user-generated evaluations quickly, publicly, and persistently, often attaching them to an identifiable individual (Aridor et al., 2024). A worker's actions can therefore generate an immediate response that remains visible to the worker, the manager, colleagues, customers, and potential future employers. This visibility distinguishes digital social pressure from a private complaint or an isolated interaction. It also raises an economic question about whether information and opinion produced by external audiences enter internal workplace decisions.

Existing work on the economics of social media primarily studies how online content changes the behavior of users, consumers, and firms. For example, experimentally reducing exposure to toxic content changes user engagement and content production (Beknazar-Yuzbashev et al., 2026). Closely related research on online reputation systems shows that consumer reviews are associated with demand, firm revenue, quality incentives, and reputation management (Chevalier and Mayzlin, 2006; Luca, 2011; Jin and Leslie, 2009; Cabral and Hortacsu, 2010; Luca and Zervas, 2016). In those settings, the principal object of evaluation is usually a product, seller, or firm. Evaluations of individual workers are economically distinct because they may be visible both to the person being evaluated and to a manager who controls access to future work. Such evaluations can therefore be related to worker behavior, managerial opportunity allocation, or both.

Professional football offers a useful setting in which to examine these relationships. Player performance is highly visible, online reactions arrive quickly, and managers repeatedly decide which players receive

opportunities to perform. Matches also occur frequently, allowing us to compare online commentary after a current match with appearance in the team’s next match and performance in a later appearance. Publicized mistakes by players such as Harry Maguire have prompted sustained online criticism and debate about continued selection (BBC Sport, 2022). Similar audience pressure arises outside sport. Following the release of *Star Wars: The Last Jedi*, director Rian Johnson faced prolonged online criticism of his creative decisions (The Guardian, 2017; IGN, 2024). Together, these examples illustrate how external audiences publicly evaluate individual workers.

We ask whether post-match online negativity predicts two distinct short-run outcomes. The first is whether a player appears in his team’s next scheduled match, an observable proxy for the allocation of a subsequent playing opportunity. This outcome reflects managerial selection together with player availability and actual participation. It is not a direct observation of the manager’s latent selection decision. The second is the player’s WhoScored rating in his next appearance, which may occur in a later fixture. The rating is observed only for players who subsequently play and is therefore conditional on the process that determines whose performance is observed. Separating these outcomes allows us to distinguish an opportunity margin from performance conditional on receiving a later opportunity.

The empirical analysis links player-match data from Europe’s five major leagues to tagged post-match Twitter¹ commentary.² We distinguish three forms of online negativity. Negative sentiment mass measures broad public dissatisfaction, toxicity mass measures hostile or uncivil language, and racial-abuse intensity measures explicitly racialized language. The measures are generated separately and are not mechanically nested. Negative sentiment and toxicity may partly proxy for publicly visible dissatisfaction with sporting performance, although tagged commentary can also concern matters unrelated to performance. Racial abuse can impose severe psychological costs while carrying less information about broad evaluations of sporting contribution. This proposed ranking concerns how different types of content map into perceived reputational costs, not whether managers regard hostile commentary as legitimate.

The results show a clear contrast between the two outcomes. Greater negative sentiment is associated with

¹We refer to the platform as Twitter throughout because our data cover 2017–2021, when the service operated under that name. It was renamed X only after our sample period.

²Prior research has used football-related Twitter data for forecasting. Brown et al. (2018) combine 13.8 million Twitter posts with contemporaneous betting prices for English Premier League matches and show that aggregate tweet tone contains information about match outcomes beyond those prices. Our analysis instead studies evaluations linked to individual players and their subsequent workplace outcomes.

a lower probability of next-team-match appearance in specifications with player and team-by-current-match fixed effects, recent-form controls, pre-match online attention, rotation measures, subsequent-fixture characteristics, and recorded injuries. Toxicity has a similar but less stable association. Racial-abuse intensity is imprecisely estimated when all three measures enter jointly. A complementary joint specification shows that positive sentiment enters with the opposite sign for appearance. By contrast, pooled associations with the rating in the player's next appearance are generally small and imprecise. This contrast is consistent with several potentially offsetting mechanisms, including managerial reputational concerns, worker stress, reputation-recovery effort, and changes in the composition of players who subsequently appear.

We also examine how realized sporting outcomes are related to the online environment faced by players. Among matches that were close in expected-goal difference, winners experience smaller pre-to-post changes in negative sentiment and toxicity than losers. Event-time estimates place the largest winner-loser differences near and shortly after kickoff. Because several pre-match coefficients also differ, the event-time patterns are best interpreted as evidence on timing rather than as causal estimates. The analysis nevertheless documents a systematic relationship between match outcomes and the online reputational environment.

The paper makes three related contributions. First, it extends research on social pressure by examining public evaluations produced outside the formal organization. Digital commentary differs from conventional workplace monitoring because it is generated by a dispersed audience, disseminated publicly, and visible to workers and managers. The analysis asks whether this external pressure predicts a managerially relevant opportunity margin as well as subsequent worker performance.

Second, the paper contributes to research on online reputation and evaluation. Most of that literature studies how reviews of firms, sellers, or products are associated with consumer demand and firm behavior. We shift attention to workers who are themselves the objects of public evaluation and to managers who allocate their subsequent opportunities. This distinction matters because an unfavorable public signal may be related to both the worker's response and the manager's willingness to continue assigning visible tasks.

Third, the paper contributes to the economics of social media by linking the content produced around

a salient event to later workplace outcomes. Professional football provides high-frequency measures of commentary, playing opportunities, and performance. It also permits separate measurement of broad negative sentiment, toxic discourse, and racialized hostility. The broadest measure of public dissatisfaction is the most stable predictor of next-team-match appearance, while the conditional-performance evidence is weaker. Taken together, the results show how external audiences may become part of the reputational environment surrounding internal opportunity allocation.

2 Conceptual Framework: Online Negativity as a Public Signal

The framework combines insights from research on social pressure and online reputation. Publicly observable judgments may be relevant to workers' behavior and to managers' allocation decisions (Bernheim, 1994; Garicano et al., 2005). Online evaluations can also create reputational incentives and be reflected in organizational choices (Chevalier and Mayzlin, 2006; Jin and Leslie, 2009; Cabral and Hortacsu, 2010; Luca, 2011; Luca and Zervas, 2016). We therefore interpret post-match online commentary as a potential public signal about an individual player. Such signals may be associated with managerial decisions through reputational concerns and with player performance through incentive and psychological channels. The framework organizes these possibilities without assuming that managers directly observe the empirical measures.

The central feature of this setting is that the signal is public. A manager may have private information about a player's fitness, tactical role, and expected contribution that is more accurate than online opinion. Even so, a highly visible public evaluation can alter the reputational environment surrounding a selection decision. At the same time, the player can observe the commentary and may respond through effort, anxiety, or distraction. The same episode of online negativity can therefore be related to an allocation decision and to performance through distinct channels. This distinction motivates the paper's separate analysis of appearance and performance conditional on a subsequent appearance.

2.1 Forms of Online Negativity and Social Pressure

Online commentary varies in intensity and form. Recent work shows that exposure to toxic content can change the behavior of users who consume it (Beknazar-Yuzbashev et al., 2026). Our question concerns whether different forms of online negativity are differentially associated with outcomes for players linked

to that content. Negative sentiment mass captures broad dissatisfaction, toxicity mass captures hostile or uncivil language, and racial-abuse intensity captures explicitly racialized language. Separate scoring procedures generate the three measures, so they can overlap without being mechanically nested.

Negative sentiment and toxicity may partly capture publicly visible dissatisfaction with sporting performance, but the tagged commentary can also concern matters unrelated to performance. Racial abuse is qualitatively different and can impose serious psychological costs. We hypothesize that it carries less information about broad performance evaluations than negative sentiment or toxicity. This distinction motivates the expected ordering of the coefficients without treating hostile content as legitimate.

The informational content of the three measures need not follow their emotional severity. Broad negative sentiment may aggregate disappointment about mistakes, effort, or match outcomes across many users. Toxic language may communicate dissatisfaction more forcefully, but it can also reflect incivility that is only weakly related to sporting contribution. Explicit racial abuse may be especially harmful to the targeted player while providing little credible information about performance. Racial discrimination also has a documented labor-market history in professional football. Using club performance and wage-bill data for English leagues from 1978–1993, Szymanski (2000) finds evidence consistent with racial discrimination in salary setting. Our racial-abuse measure captures a distinct and much more recent phenomenon, racialized online discourse, but this historical evidence underscores why racial hostility in football can have labor-market relevance beyond ordinary performance criticism. Managers may consequently discount different forms of commentary at different rates. Players may also experience different psychological responses to them. The empirical coefficients reflect these combined relationships and should not be interpreted as a ranking of the social harm caused by each form of language.

These considerations motivate three empirical hypotheses. First, greater online negativity may predict a lower probability of next-team-match appearance. Second, negative sentiment mass and toxicity mass may be more closely related to this appearance outcome than racial-abuse intensity. Third, in joint specifications, racial-abuse intensity may add little predictive power after the other measures enter. The ranking requires measure-specific mappings from online content to perceived reputational costs. It does not follow mechanically from the generic model below.

2.2 Managerial Selection Under Public Reputation Signals

Online commentary may create reputational pressure for managers because their selection decisions are publicly scrutinized. Even when managers possess superior private information about player quality, selecting a heavily criticized player may expose them to costs among supporters, journalists, club executives, and other stakeholders. This mechanism is related to models in which public social judgments enter decision-makers' reputational concerns (Bernheim, 1994).

Let $s_{it} \geq 0$ denote the observed intensity of online negativity linked to player i following match t . Managers need not respond mechanically to the volume of commentary. Instead, the commentary may contribute to a latent *reputational-adversity signal*, denoted by q_{it} , where

$$q_{it} = Q(s_{it}), \quad Q'(s_{it}) > 0.$$

The function $Q(\cdot)$ maps observed online negativity into reputational adversity. Higher q_{it} represents a worse public evaluation. The mapping need not be one-to-one because managers may place different weights on different forms of commentary and may discount content they regard as unrepresentative or abusive.

Managers care about both team performance and their own reputation. Let the manager's utility from selecting player i for the subsequent match be

$$U_{it} = \lambda C_{it} - (1 - \lambda)R_{it} + \varepsilon_{it},$$

where C_{it} denotes the player's expected sporting contribution, R_{it} denotes the reputational cost associated with selecting the player, ε_{it} is an idiosyncratic shock, and $\lambda \in [0, 1]$ measures the relative weight the manager places on sporting performance versus reputational considerations. Higher values of λ indicate that the manager places greater emphasis on expected sporting contribution, whereas lower values imply that reputational concerns play a larger role in the selection decision.

The reputational cost depends on the player's reputational-adversity signal:

$$R_{it} = R(q_{it}) = R(Q(s_{it})),$$

with $R'(q_{it}) > 0$. The model therefore allows greater supporter dissatisfaction, media scrutiny, or organizational pressure when a manager selects a player with a more adverse public evaluation.

Normalize the manager's utility from not selecting player i to zero. Let $d_{i,t+1} = 1$ denote the latent decision to select the player for the team's next match. The manager selects the player whenever

$$U_{it} \geq 0.$$

Under the maintained structure, a higher perceived reputational cost lowers the probability of latent selection:

$$\frac{\partial \Pr(d_{i,t+1} = 1)}{\partial q_{it}} < 0,$$

and, because $q_{it} = Q(s_{it})$,

$$\frac{\partial \Pr(d_{i,t+1} = 1)}{\partial s_{it}} = \frac{\partial \Pr(d_{i,t+1} = 1)}{\partial q_{it}} Q'(s_{it}) < 0.$$

Thus, provided that greater online negativity raises perceived reputational adversity, the model predicts a lower probability of latent selection.

The observed outcome, denoted by $y_{i,t+1}$, equals one when the player appears in the team's next scheduled match. It is an observable proxy for a realized playing opportunity, not a direct measure of $d_{i,t+1}$. Appearance reflects the latent managerial decision together with availability and actual participation. The empirical analysis therefore examines whether the three online-negativity measures predict this observed outcome.

The model is deliberately parsimonious. It does not require a manager to follow supporter opinion or to sacrifice expected sporting performance whenever criticism rises. A manager who places a high weight on expected contribution, or who regards a particular form of commentary as uninformative, may change the selection threshold little. Reputational considerations are more likely to matter when expected contributions are similar across available players or when continued selection of a criticized player is

especially visible. The empirical appearance coefficient is therefore an average conditional association across decisions made under different tactical, reputational, and availability conditions.

2.3 Performance Conditional on a Subsequent Appearance

Performance is observed only when a player subsequently appears. Let $a(i, t)$ denote player i 's next appearance after match t , which need not be the team's next scheduled match. Realized performance in that appearance is

$$p_{i,a(i,t)} = F(\theta_i, e_{i,a(i,t)}, m_{i,a(i,t)}, X_{i,a(i,t)}),$$

where θ_i denotes underlying ability, $e_{i,a(i,t)}$ denotes effort, $m_{i,a(i,t)}$ denotes psychological burden, and $X_{i,a(i,t)}$ denotes factors specific to that appearance. The observed conditional association between online negativity and this performance outcome may combine two behavioral channels with a change in the composition of players whose ratings are observed.

2.3.1 Reputation-Recovery and Career-Concern Channel

A more adverse public evaluation may strengthen a player's incentive to restore his standing with managers, supporters, and prospective employers. One possible response is greater effort in the next appearance. Let

$$e_{i,a(i,t)} = e(q_{it}), \quad e'(q_{it}) > 0.$$

If greater effort is associated with better performance, then this channel works in a positive direction. The mechanism may be stronger for younger players, who generally have shorter career histories and greater uncertainty about future contracts. This age pattern is a hypothesis because the analysis does not observe effort or reputation recovery directly.

The economic intuition is related to career concerns. Public criticism may lower a player's perceived standing with the current club, supporters, or prospective employers. A player who expects future selection, contract terms, or transfer opportunities to depend on subsequent performances may respond by exerting more effort when another opportunity arrives (Holmström, 1999; Dewatripont et al., 1999). This reasoning does not imply that effort necessarily rises or that greater effort always offsets other

burdens associated with online negativity.

2.3.2 Stress Channel

Online negativity may also be associated with greater anxiety, stress, or distraction. Let

$$m_{i,a(i,t)} = m(s_{it}), \quad m'(s_{it}) > 0.$$

If greater psychological burden is associated with worse performance,

$$\frac{\partial p_{i,a(i,t)}}{\partial m_{i,a(i,t)}} < 0,$$

then this channel works in a negative direction. The empirical analysis does not observe psychological burden, so it cannot isolate this mechanism from other explanations.

Unlike the reputation-recovery channel, the stress channel can operate even when the commentary carries no useful information about sporting ability. Persistent, personalized, or abusive content may interfere with concentration, sleep, preparation, and confidence before the next appearance. Experimental evidence shows that exposure to rude treatment can impair task performance through disruptions to cognitive processes (Porath and Erez, 2007). These responses may also differ across players and types of commentary. The available data contain performance outcomes rather than direct measures of psychological well-being, which limits the paper to describing the net conditional relationship.

2.3.3 Contribution Selection

The managerial selection mechanism in Section 2.2 may also affect the composition of players whose subsequent performance is observed. If greater online negativity raises the reputational cost of selecting a player, a criticized player may need a sufficiently high expected sporting contribution to remain worth selecting. Thus, conditional on making a subsequent appearance, players facing greater online negativity may be positively selected on expected contribution. More formally, under the additional assumption that selection is monotone in expected sporting contribution, managers may continue to select heavily criticized players only when those players have sufficiently high expected contributions. The conditional expectation

$$E[C_{i,a(i,t)} \mid y_{i,a(i,t)} = 1, s_{it}]$$

may then rise with online negativity. This positive selection does not follow from the threshold model alone. It requires assumptions about the joint distribution of expected contribution, online negativity, and the idiosyncratic selection shock. We therefore treat it as a possible source of non-random composition rather than a formally identified component of the observed performance relationship.

This composition channel is important because a rating is observed only after a player returns to the field. If adverse public evaluation is associated with a higher threshold for receiving another opportunity, the players whose later ratings remain observable may differ from those who do not appear. Conversely, online attention may be concentrated among prominent players who are likely to play regardless of criticism. Either pattern can change the relationship between online negativity and observed ratings without representing a behavioral response by the player.

2.4 Implications for Empirical Estimation

The observed association between online negativity and performance in the player’s next appearance can combine offsetting behavioral responses and changes in whose performance is observed. Reputation-recovery incentives may be associated with greater effort, stress may be associated with worse performance, and non-random subsequent appearance may change the composition of the rating sample. The overall conditional relationship is therefore ambiguous, and the regressions do not separately estimate these channels.

Regressions without player fixed effects can also reflect persistent cross-player differences in ability, prominence, baseline attention, or managerial trust. Adding player fixed effects removes stable additive differences among the Twitter-linked players. It does not eliminate time-varying selection into the rating sample or make account linkage random. Comparing specifications with and without player fixed effects therefore shows whether persistent cross-player heterogeneity matters for the coefficient. A downward shift may be consistent with positive sorting on ability or prominence, but it is not a direct test of managerial selection.

The framework also motivates heterogeneity in the empirical relationships. Younger players may have

stronger career concerns and a greater incentive to restore their public standing after adverse evaluations. Older players may differ in reputation, role, or exposure. Because effort, stress, and managerial expectations are not observed, subgroup patterns can inform the relevance of these channels without isolating a single mechanism.

3 Data

The analysis uses a player-match panel covering the Bundesliga, La Liga, Ligue 1, the Premier League, and Serie A from the 2017–2018 through 2020–2021 seasons. The sample combines data from multiple sources in the project repository, including WhoScored player-match ratings and event files, FBRef match and roster files, and post-match Twitter windows containing tagged tweets. The sample period is determined by the availability of comparable WhoScored league-player files across all five leagues through the 2020–2021 season.

The empirical analysis focuses on players with identifiable Twitter accounts in the FBRef roster files because the negativity measures rely on tagged social-media exposure. The richest appearance specification uses 38,653 player-match observations. The rating sample is smaller, with 33,512 observations, because the outcome is observed only when the player makes a subsequent appearance within 14 days. Table 1 reports counts and shares for the league, age, and nationality splits in the main results. Appendix Table A2 reports broader linked-panel and Twitter-account sample statistics.

The first dependent variable is next-team-match appearance, an observable proxy for a subsequent playing opportunity. It equals one if the player appears in his team’s next scheduled match and zero if the team plays but the player does not appear. The outcome reflects managerial selection together with availability and actual participation, so it is not a direct measure of the manager’s latent decision. The appearance sample restricts attention to next-team matches that occur within 14 days of the current match.

The second dependent variable is the player’s WhoScored rating in his next appearance. For each player-match observation, we recover the final rating from the player-actions files and assign the next observed rating chronologically. WhoScored.com constructs this performance measure using a proprietary algorithm that combines more than 200 match statistics. These include offensive actions such as goals, assists, shots, key passes, and successful dribbles, defensive actions such as tackles, interceptions, clearances, and aerial duels, passing statistics, and negative events such as fouls, missed chances, defen-

sive errors, and disciplinary actions (WhoScored.com, 2026). As in the appearance analysis, the rating sample is restricted to cases in which the player's next appearance occurs within 14 days of the current match.

Online negativity is measured using tweets posted within 24 hours after a match that contain the recorded player, team, or match-related tags. The analysis links a tweet to a player-match observation through these tags. This construction does not establish that every tweet is addressed to, or specifically about, the player. We therefore describe the commentary as linked to the player-match unless the construction supports a narrower statement. Restricting the analysis to the post-match period captures immediate public reactions while limiting unrelated discussion. Tweets are translated into English so that a common set of language-processing tools can be applied. The Twitter pipeline then generates three measures, each designed to capture a distinct dimension of post-match online commentary.

The first measure is negative sentiment mass. Sentiment analysis is a natural language processing (NLP) technique that estimates the emotional tone expressed in text (Liu, 2012). In this study, sentiment is measured using the Valence Aware Dictionary and sEntiment Reasoner (VADER), a rule-based sentiment analysis tool developed specifically for social media text (Hutto and Gilbert, 2014). VADER assigns positive, neutral, negative, and composite sentiment scores to each tweet. The *negative sentiment mass* is calculated as the sum of the negative sentiment scores across all tagged tweets posted after a match. This measure reflects the overall volume and intensity of negative opinions or dissatisfaction expressed online, regardless of whether the language is abusive or offensive. For the complementary analysis, positive sentiment mass is defined analogously by summing the positive VADER component across the same tagged post-match tweets.

The second measure is toxicity mass. Toxicity refers to language that is rude, insulting, aggressive, or likely to discourage constructive online discussion. Rather than measuring general negativity, toxicity focuses specifically on hostile or uncivil communication. The repository estimates toxicity using a lexical toxicity proxy, which identifies words and expressions associated with abusive or offensive language and assigns each tweet a toxicity score. The *toxicity mass* is obtained by summing these tweet-level toxicity scores over all tagged post-match tweets. Higher values indicate a greater amount of hostile discourse following a match. Similar approaches to toxicity detection have been widely adopted in computational social science and online content moderation research (Wulczyn et al., 2017).

The third measure is racial-abuse intensity. Unlike general toxicity, this measure focuses on language concerning individuals or groups based on race or ethnicity. The repository applies a conservative detector designed to identify explicit racial slurs, stereotypes, and other forms of racially abusive language while limiting false positives. Tweet-level racial-abuse scores are summed across tagged post-match tweets to produce *racial-abuse intensity*. The variable captures explicitly racialized language in the tagged post-match conversation. Research on automated hate-speech detection similarly distinguishes racial abuse from general offensive language because the two phenomena are related but conceptually distinct (Davidson et al., 2017).

The three measures capture different dimensions of online discourse. Negative sentiment mass measures overall emotional negativity, toxicity mass measures hostile or uncivil language, and racial-abuse intensity measures explicitly race-related abusive content. Because each measure uses a different scoring procedure, they are not mechanically nested. A tweet expressing disappointment may have a high negative-sentiment score but no toxicity or racial-abuse score. Conversely, racially abusive language may receive high toxicity and racial-abuse scores even if its overall sentiment score is not extremely negative. Treating the measures separately distinguishes general criticism, abusive behavior, and explicitly racialized hostility.

Because the post-match negativity masses are highly right-skewed and contain many zeros, the appearance and performance specifications use

$$\log(1 + \text{negativity mass}).$$

Throughout the paper, $\log$ denotes the natural logarithm.

The transformation reduces the influence of extreme realizations while retaining observations with zero measured negativity. In the main Twitter-only sample, 86.3% of observations contain at least one tagged post-match tweet within the 24-hour window. The mean negative sentiment mass is 6.37, the mean toxicity mass is 1.05, and the mean racial-abuse intensity is 0.089. The distributions contain substantial mass at zero, especially for racial-abuse intensity, further motivating the log transformation. This transformation does not apply to the match-outcome analysis described below. In that analysis, the dependent variables are changes in the underlying negativity scores between the 24-hour pre-match and

post-match windows, measured on their original scale.

Table 2 reports descriptive outcome means after dividing the 33,512-observation rating sample into observations above and at or below the sample mean of each online-negativity measure. The appearance column is therefore calculated within the rating sample and does not describe the full appearance sample. High-exposure observations have higher subsequent ratings for all three measures, while appearance probabilities differ little between the high- and low-exposure groups. These comparisons are descriptive because highly exposed observations may involve more prominent players, unusual match contexts, or other characteristics that also predict subsequent outcomes.

The regressions additionally include a rich set of controls designed to account for current match conditions, player characteristics, baseline online exposure, prior form, next-match scheduling, and outcome-aligned injury availability. Current-match controls include match-result indicators, home status, goal difference, expected-goal differential ($xGD = xG - xGA$), and a wage-difference proxy constructed from the FBRef match files. Player controls include age, broad playing position, and detailed WhoScored role indicators. The main appearance specifications also include rotation controls based on the player's most recent appearance, rather than the team's immediately preceding match, through indicators for whether the player started or appeared as a substitute in that appearance. The regressions further include Transfermarkt injury controls aligned to the date of the team's next match. Appendix Table A1 summarizes Transfermarkt-observed injury shares by league and season in the appearance sample.

To account for baseline social-media exposure and pre-existing sentiment, the specifications also include pre-match tagged-tweet volume and pre-match negativity measures constructed over the 24 hours before kickoff. Prior-form controls include the player's rating in his previous appearance, the rolling average of the previous three ratings, his career-average rating through the prior appearance, and days since that appearance. Next-match schedule controls include days until the next match, next-match home status, and the next-match wage-difference proxy.

The results section further considers heterogeneity across league groups, age groups, and nationality status. The league split compares Premier League observations with observations from the other four major European leagues. The age split separates players aged 26 or younger from players older than 26. The split appearance and rating tables use a domestic-versus-foreign nationality split relative to the

league country.

4 Methodology

4.1 Online Negativity, Appearance, and Performance

The empirical strategy relates post-match online negativity in current match m to two subsequent outcomes defined at different events. Let $Y_{im}^{(A)}$ be an indicator equal to one if player i appears in his team's next scheduled match. Let $Y_{im}^{(R)}$ be the player's WhoScored rating in his next appearance, which may occur in a later fixture. Our main specification is

$$Y_{im}^{(k)} = \beta^{(k)} P_{im}^{(g)} + X_{im}' \gamma^{(k)} + Z_{im}^{(k)'} \delta^{(k)} + \alpha_i + \psi_{j(i,m),m} + \eta_{im}^{(k)}, \quad k \in \{A, R\},$$

where $P_{im}^{(g)}$ is the log-transformed post-match negativity measure of type g , X_{im} contains current-match observables, player characteristics, pre-match Twitter exposure, and prior form, and $Z_{im}^{(k)}$ contains controls aligned with the relevant subsequent event. The appearance equation is estimated as a linear probability model. Player fixed effects, α_i , absorb stable additive differences across players. Team-by-current-match fixed effects, $\psi_{j(i,m),m}$, absorb circumstances shared by teammates in the same current match. The benchmark coefficient therefore compares teammates in the same current match and relates each player's subsequent outcome to deviations in online negativity from his typical level. $\beta^{(A)}$ describes the conditional association with next-team-match appearance, and $\beta^{(R)}$ describes the corresponding association with the rating in the player's next appearance.

A key empirical challenge is that post-match online negativity may be correlated with unobserved determinants of subsequent appearance or performance. The benchmark comparison is within a player over time and within a team in the same current match. In plain language, it asks whether two teammates who share the same current-match environment have different subsequent outcomes when one receives more post-match negativity than is usual for him, conditional on the observed covariates. This comparison removes many persistent player differences and common team-match circumstances, but it does not make online negativity randomly assigned.

One concern is selection into the Twitter-linked sample. The negativity measures rely on tweets associated with identifiable player accounts, and maintaining such an account is not randomly determined.

Players with identifiable accounts may differ from other professional footballers in prominence, online engagement, sensitivity to commentary, or other unobserved attributes. The estimates therefore describe conditional relationships among the Twitter-linked population and need not generalize to players without identifiable accounts. Player fixed effects absorb stable additive differences within this population, but account linkage and external validity remain potential limitations.

A second concern is that time-varying off-field events may predict both online negativity and subsequent football outcomes. Legal or disciplinary proceedings, transfer disputes, contract negotiations, or conflicts with managers and teammates may increase negative attention while independently changing a player's concentration, availability, or on-field performance. An event that begins before current match m and persists afterward could generate an association even if the post-match commentary itself plays no role in the subsequent outcome. Pre-match online measures and prior-form controls address observable traces of ongoing events, but sudden player-specific shocks can remain.

Strategic squad rotation creates a related challenge for the appearance outcome. Clubs competing in domestic and European competitions often face congested schedules and may rest important players around high-priority fixtures. A player could perform poorly and receive greater negativity in current match m , then miss the team's next match because the manager had already planned to use him in another competition. The observed non-appearance would then reflect a rotation decision rather than a response to commentary. Tactical preferences, accumulated workload, travel, recovery time, and the importance of the next fixture can generate similar patterns.

Injuries and other fitness limitations can also be associated with both the explanatory variables and outcomes. A player carrying an injury may perform poorly, receive more negative commentary, and subsequently be unavailable. A less severe condition may not prevent a later appearance but may be associated with a lower rating when the player returns. Recorded injury measures address part of this concern. Minor injuries, fatigue, and fitness limitations that are not reported or are inaccurately dated may still contribute to residual confounding.

The benchmark specification combines player fixed effects with team-by-current-match fixed effects to address important components of these concerns. Player fixed effects absorb stable additive differences among Twitter-linked players, including persistent differences in ability, popularity, baseline online

exposure, managerial trust, average injury propensity, and typical appearance or performance. The estimating variation therefore comes from whether a given player receives more or less negativity than he typically receives. Player fixed effects do not absorb player-specific time-varying shocks, allow every player to have a different response to negativity, or resolve non-random inclusion in the linked sample.

Team-by-current-match fixed effects compare teammates in the same current match. They absorb circumstances shared by those teammates, including the manager, opponent, match importance, team performance, tactical environment, lineup conditions, and contemporaneous team shocks. They also absorb the common component of fixture congestion and current-match context faced by the team. The coefficient is estimated from differences in post-match negativity across teammates after accounting for each player's average outcome and the observed player-specific covariates. Standard errors are clustered at the player level throughout.

The remaining controls address observable time-varying factors that are not absorbed by the fixed effects. Current-match controls include the realized result, home status, goal difference, expected-goal differential, and the wage-difference proxy in specifications where those variables are not collinear with the fixed effects. These measures account for sporting circumstances that may be associated with both the tone of commentary and later outcomes. In the benchmark team-by-current-match fixed-effects models, team-level current-match variables common to all teammates are absorbed rather than separately estimated.

Pre-match Twitter controls include tagged-tweet volume and pre-match online-negativity measures constructed over the 24 hours before kickoff. They account for baseline attention and help distinguish reactions following current match m from criticism or controversy already present beforehand. Prior-form controls include the player's rating in his previous appearance, the rolling average of his previous three ratings, his career-average rating through the prior appearance, and days since that appearance. These variables capture recent performance and availability trajectories that can accompany persistent off-field events, deteriorating form, or emerging fitness problems.

Schedule and rotation controls address player-specific rotation risk not absorbed by team-by-current-match fixed effects. Days until the team's next match measure recovery time. Next-match home status

and the wage-difference proxy describe observable features of the subsequent fixture. Indicators for whether the player started or entered as a substitute in his previous appearance, together with days since that appearance, capture recent workload and role. Player fixed effects also absorb persistent differences in squad status, such as whether a player is usually a regular starter, rotation option, or fringe player.

Transfermarkt injury measures are aligned to the date relevant for each outcome. For next-team-match appearance, injury status is aligned with the team’s next scheduled match. For the rating outcome, it is aligned with the player’s next appearance. This distinction helps separate recorded unavailability from performance after a player returns. Prior-form measures and time since the previous appearance provide additional information about interrupted availability. These controls cannot capture every minor or undisclosed condition.

Taken together, the fixed effects and control families substantially narrow the set of plausible confounding explanations. They cannot eliminate selection into the Twitter-linked population, sudden player-specific off-field shocks, unobserved rotation plans, or unreported fitness limitations, nor do they reveal whether a manager observed the relevant commentary. We therefore interpret $\beta^{(A)}$ and $\beta^{(R)}$ as conditional within-player relationships among Twitter-linked players rather than as causal effects.

4.2 Match Outcomes and Online Negativity

The main analysis studies conditional associations between post-match online negativity and subsequent player outcomes. This subsection instead describes how online negativity differs with realized match outcomes. The online-negativity outcomes used here differ from the post-match negativity-mass variables in the appearance and performance regressions.

Let $B_{im,pre}^{(g)}$ and $B_{im,post}^{(g)}$ denote the online-negativity score of type g linked to player i during the 24 hours before and after match m , respectively. The index g denotes negative sentiment, toxicity, or racial abuse. We define the pre-to-post-match change as

$$\Delta B_{im}^{(g)} = B_{im,post}^{(g)} - B_{im,pre}^{(g)}. \quad (1)$$

These outcomes remain on their original score scales and are not transformed using $\log(1 + x)$.

To improve winner–loser comparability, we focus on matches that were close in expected-goal difference.

From the perspective of player i 's team,

$$xGD_{im} = xG_{im} - xGA_{im}, \quad (2)$$

where xG_{im} and xGA_{im} are expected goals and expected goals against in match m . Values near zero indicate that the teams generated similarly valuable scoring opportunities, even when the realized outcomes differed. Within a symmetric bandwidth around $xGD_{im} = 0$, we estimate the close-match local linear comparison

$$\begin{aligned} \Delta B_{im}^{(g)} = & \beta_g \text{Win}_{im} + \pi_{1g} xGD_{im} \\ & + \pi_{2g} (\text{Win}_{im} \times xGD_{im}) + X'_{im} \Gamma_g + \alpha_i + \mu_m + \varepsilon_{im}^{(g)}, \end{aligned} \quad (3)$$

where Win_{im} equals one if the player's team won and zero if it lost. Draws are excluded. The vector X_{im} contains player- and team-specific covariates, α_i denotes player fixed effects, and μ_m denotes match fixed effects.

Player fixed effects absorb persistent differences across players in online exposure and the propensity to receive negative commentary. Match fixed effects absorb characteristics shared by both teams in the same match.³ Players on opposing teams in the same match have different values of Win_{im} and opposite-signed values of xGD_{im} . The win-by- xGD interaction allows the association with xGD to differ by realized match outcome. The coefficient β_g is the conditional winner–loser difference in the pre-to-post change at $xGD_{im} = 0$. Winning is not assigned by crossing an xGD threshold, so the specification is an xGD -bandwidth comparison.

We estimate Equation (3) within $|xGD_{im}| \leq 0.25$ and $|xGD_{im}| \leq 0.50$. For each bandwidth, we report an unweighted specification and specifications weighted by either total tweets associated with the player-match observation or tweets in the post-match window. Standard errors are clustered at the match level.

As a robustness exercise, we construct matched samples of winning and losing player-match observations with similar values of xGD_{im} . The alternatives vary the maximum permissible xGD distance and whether

³We use match fixed effects rather than team-by-current-match fixed effects in this analysis. Match outcome and xGD_{im} are constant across players from the same team in a given match and would be collinear with team-by-current-match fixed effects. Match fixed effects preserve the within-match, between-team variation needed for the winner–loser comparison.

losing observations may be reused as matches. Equation (3) is then re-estimated in each matched sample. This exercise assesses whether the estimates are sensitive to expected-goal overlap between winning and losing observations.

To examine timing, we complement the close-match analysis with an event-time specification. Let $B_{imh}^{(g)}$ denote the online-negativity score of type g linked to player i in event hour h around match m . The hourly outcomes remain on their original score scales. We estimate

$$B_{imh}^{(g)} = \delta_g \text{Win}_{im} + \sum_{\substack{h=-24 \\ h \neq -1}}^{24} \theta_{hg} [\text{Win}_{im} \times \mathbf{1}\{H_{imh} = h\}] + \alpha_i + \mu_m + \lambda_h + \varepsilon_{imh}^{(g)}, \quad (4)$$

where H_{imh} is event time relative to kickoff. The hour immediately before kickoff, $h = -1$, is omitted. Player fixed effects, α_i , absorb persistent differences across players in online negativity. Match fixed effects, μ_m , absorb characteristics shared by both teams in the same match. Event-hour fixed effects, λ_h , absorb common changes in online negativity at each event hour. The main win term δ_g captures the winner–loser difference in the omitted pre-match hour. The interaction coefficients θ_{hg} capture winner–loser differences at each event hour relative to that omitted hour. Identification therefore comes from how the within-match difference between players on winning and losing teams changes across event hours relative to hour -1 , after accounting for persistent player differences and common match and event-hour conditions.

5 Results

5.1 Online Negativity and Next-Team-Match Appearance

Table 3 reports conditional associations between post-match online negativity and next-team-match appearance. In the single-measure specifications, a one-unit increase in $\log(1 + \text{negative-sentiment mass})$ is associated with a 2.4-percentage-point lower appearance probability (s.e. 0.5 percentage points). A one-unit increase multiplies $1 + \text{negative-sentiment mass}$ by e . The corresponding coefficients are -0.020 for toxicity mass and -0.017 for racial-abuse intensity. In the joint specification, the negative-sentiment coefficient remains -0.021 , the toxicity coefficient is -0.013 , and the racial-abuse coefficient is close to zero and imprecisely estimated. This ordering is consistent with the proposed public-evaluation

mechanism. The comparatively weak racial-abuse coefficient concerns its predictive content for short-run appearance rather than the broader social or labor-market importance of racial hostility.

The single-measure estimates describe the combined relationship between each measure and appearance, including the online content that it shares with the other measures. The joint specification instead compares their conditional associations while holding the other two measures fixed. The attenuation of the toxicity coefficient and the near-zero racial-abuse coefficient indicate that some of their single-measure relationships overlap with broader negative sentiment. Negative sentiment therefore remains the most stable predictor of the observed playing-opportunity margin.

Although the main analysis focuses on negative commentary, social pressure can also operate through approval. Favorable evaluations may reinforce a player's reputation and increase the perceived benefit of continued selection. We therefore use positive sentiment as a complementary test of the public-evaluation mechanism. If this mechanism reflects the broader reputational environment surrounding managerial decisions, positive sentiment should predict appearance in the opposite direction to negative sentiment.

Appendix Table A3 reports this test. Positive sentiment is positive but imprecisely estimated in the single-measure appearance specification. In the four-measure joint specification, which includes negative sentiment, positive sentiment, toxicity, and racial-abuse intensity, its coefficient is 0.017 and statistically significant at the 1 percent level. Negative sentiment remains negative in the same regression. These opposite signs complement the public-evaluation interpretation. Favorable evaluations are associated with a higher subsequent appearance probability, whereas unfavorable evaluations are associated with a lower one.

Panel B reports age-group estimates. In the joint specifications, the negative-sentiment coefficient is -0.020 for players aged 26 or younger and -0.024 for older players. The toxicity point estimate is -0.018 for younger players and -0.009 for older players. Racial-abuse estimates are close to zero in both groups. Because the table does not report cross-group equality tests, the subgroup patterns should be interpreted descriptively rather than inferred from differences in statistical significance.

Table 4 reports descriptive estimates by league group and nationality. In the joint specification, the negative-sentiment coefficient is negative but imprecisely estimated for Premier League players (-0.014 ,

s.e. 0.011) and is -0.014 (s.e. 0.006) in the other four leagues. The corresponding toxicity estimates are -0.018 and -0.026 . Racial-abuse intensity is imprecisely estimated in both league groups.

In the nationality split, the joint negative-sentiment coefficient is -0.030 for domestic players and -0.012 for foreign players. The toxicity coefficient is close to zero for domestic players and -0.025 for foreign players. The racial-abuse coefficient for foreign players changes from -0.031 in the single-measure specification to -0.012 in the joint specification.

Across the reported subgroups, appearance coefficients are most consistently negative for negative sentiment, while toxicity estimates are less uniform. The results point to next-team-match appearance as the outcome most consistently associated with online negativity.

The subgroup patterns also help describe where the pooled association originates. Negative-sentiment estimates remain negative across the reported age, league, and nationality splits, while the relative magnitudes of toxicity and racial-abuse estimates vary more. This broad pattern is compatible with negative sentiment capturing a more general public evaluation and the other measures capturing narrower dimensions of discourse. Differences in player roles, baseline attention, match contexts, or measurement precision may account for part of the variation across subgroups.

5.2 Online Negativity and Performance Conditional on Appearance

Tables 5 and 6 examine the player's WhoScored rating in his next appearance. These estimates are conditional on a subsequent appearance and may reflect offsetting differences in stress, effort, and the composition of players who continue to play.

Panel A of Table 5 presents the preferred specification with player and team-by-current-match fixed effects. The negative-sentiment coefficient is -0.015 in the single-measure specification and -0.013 in the joint specification. The toxicity and racial-abuse estimates are also small and imprecisely estimated. In contrast to the appearance results, the pooled estimates provide limited evidence of a stable relationship between online negativity and the rating in the player's next appearance.

The positive-sentiment results in Appendix Table A3 reinforce this contrast. Its coefficient is close to zero in both the single-measure and four-measure joint rating specifications. Positive sentiment thus complements the appearance result without revealing a corresponding pooled relationship with

subsequent performance.

Panel B omits player fixed effects while retaining team-by-current-match fixed effects and the other controls. The negative-sentiment coefficient is 0.009 in the single-measure specification and 0.007 in the joint specification. Adding player fixed effects in Panel A changes these estimates to -0.015 and -0.013 , respectively. This shift indicates that persistent cross-player heterogeneity matters for the estimated relationship rather than serving as a direct test of managerial selection.

One plausible source of the difference is that players who routinely attract more online attention also differ persistently in ability, prominence, or expected playing time. Those cross-player differences can generate a different unconditional pattern from the within-player relationship. The player fixed-effects estimates remove stable additive differences of this kind, but they remain conditional on making a later appearance. They therefore cannot reveal the ratings that would have been recorded for players who did not return within the sample window.

Panel C reports results by age. For players aged 26 or younger, the negative-sentiment coefficient is -0.007 in the single-measure specification and -0.009 in the joint specification. For older players, the corresponding estimates are -0.036 and -0.037 . The joint racial-abuse coefficient for older players is positive at 0.066, while toxicity is imprecisely estimated. These patterns are consistent with several possible differences across players, including reputation, role, and exposure, but do not distinguish among those channels.

Table 6 reports additional patterns by league group and nationality. In the Premier League joint specification, the negative-sentiment coefficient is -0.055 and the toxicity coefficient is 0.035. In the other leagues, the corresponding estimates are -0.010 and -0.044 . None of the three measures is precisely estimated for either domestic or foreign players. The same qualification about formal cross-group tests applies to these estimates. Overall, the conditional-performance estimates are less stable than the appearance estimates.

5.3 Match Outcomes and Online Negativity

Table 7 reports winner–loser differences in pre-to-post changes in online negativity among matches that were close in expected-goal difference. Across both bandwidths and all three weighting schemes, winning is associated with a smaller pre-to-post change in negative sentiment and toxicity. At the 0.50 bandwidth,

the negative-sentiment coefficient ranges from -0.020 in the unweighted specification to -0.021 with post-match tweet weights. The corresponding toxicity estimates range from -0.002 to -0.003 . The racial-abuse estimates are much smaller and less stable, ranging from 0.0001 to -0.0001 at the 0.50 bandwidth. These estimates summarize conditional winner–loser differences near $xGD = 0$.

Appendix Table A4 shows a similar descriptive pattern in matched winner–loser samples. The negative-sentiment coefficients range from -0.0165 to -0.0183 , and the toxicity coefficients are negative in all four matched samples. Racial-abuse estimates remain smaller and less stable. Together, the comparisons indicate that realized match outcomes are systematically related to the subsequent online reputational environment, especially for broad negative sentiment and toxicity.

The alternative bandwidths change how closely the sample is concentrated around balanced expected-goal performance. The weighting schemes give more influence to player-match observations supported by more tagged commentary. The similar negative-sentiment and toxicity patterns across these choices indicate that the result is not confined to a single bandwidth or weighting rule. The matched comparisons provide a further check on overlap in expected-goal difference. The consistency across these checks strengthens the evidence that realized match outcomes are related to subsequent online negativity.

Figure 1 describes how winner–loser differences in online negativity evolve around kickoff. Each coefficient measures the winner–loser difference at the indicated event hour relative to the omitted hour immediately before kickoff, $h = -1$. For negative sentiment, the estimate is -0.0135 at hour 0, -0.0248 at hour 1, and -0.0391 at hour 2. It remains negative at hours 6 and 12, with estimates of -0.0255 and -0.0305 , and is small and imprecisely estimated at hour 24. Toxicity is -0.0015 at hour 0, -0.0047 at hour 1, and -0.0028 at hour 2. It is close to zero at hours 6 and 12 and positive at hour 24. The racial-abuse estimates are smaller and do not display a stable pattern.

Several pre-match coefficients are also precisely estimated. Negative sentiment differs at hours -12 , -6 , and -2 , toxicity differs at those same hours, and racial-abuse intensity differs at hour -2 . These pre-match differences make the figure most informative about timing rather than causality. Appendix Table A5 reports the complete estimates and standard errors.

6 Conclusion

Digital platforms have expanded the scale, speed, and visibility of public evaluations directed at individual workers. Professional football provides a setting in which those evaluations can be linked to high-frequency measures of playing opportunities and subsequent performance. The analysis distinguishes broad negative sentiment, toxic language, and explicitly racialized hostility rather than treating all unfavorable online content as a single form of pressure.

The main empirical contrast is between next-team-match appearance and performance conditional on a subsequent appearance. Greater post-match online negativity, especially negative sentiment, is associated with a lower probability of appearing in the team's next match. Next-team-match appearance is an observable proxy for a subsequent playing opportunity, not a direct measure of the manager's latent selection decision. By contrast, the pooled relationships between online negativity and the rating in the player's next appearance are generally small and imprecisely estimated.

The specifications compare players with themselves over time and, in the benchmark models, compare teammates in the same current match. The appearance association remains after accounting for recent performance, pre-match online attention, rotation, subsequent-match characteristics, and recorded injuries. Toxicity is also negatively associated with appearance in several specifications, while racial-abuse intensity is less stable in the joint models. In a complementary joint specification, positive sentiment enters with the opposite sign to negative sentiment for appearance. The subgroup estimates show broadly consistent negative-sentiment patterns, although the tables do not report formal cross-group equality tests.

The contrast across measures is economically informative, but it concerns their informational content rather than the social harm they cause. Negative sentiment may provide a broader signal of public dissatisfaction, while toxicity combines evaluation with hostile expression. Racial abuse may impose severe psychological costs while conveying little credible information about sporting contribution. The weaker conditional-performance evidence is also compatible with offsetting processes. Stress may be associated with worse performance, reputation-recovery incentives may be associated with greater effort, and the set of players who subsequently appear may be non-random. The available data do not separate these mechanisms.

Close-match comparisons provide complementary evidence on the online environment that follows sporting outcomes. Winning is associated with a smaller pre-to-post change in negative sentiment and toxicity than losing among matches with similar expected-goal differences. Event-time estimates place the largest winner–loser differences near and shortly after kickoff, while the presence of pre-match differences makes these estimates most informative about timing. Changes in racial-abuse intensity are smaller and less stable.

Several limitations qualify these findings. Players with identifiable Twitter accounts may differ from other professional footballers, and player-specific time-varying shocks, unobserved rotation plans, or unreported fitness problems may remain. The estimates therefore represent conditional within-player relationships. Further research using exogenous variation in access to, visibility of, or amplification by digital platforms could help distinguish managerial responses from worker responses.

More broadly, the findings show how public evaluations produced outside an organization can become part of the reputational environment in which managers allocate opportunities and workers perform. This possibility extends beyond sport to other visible labor markets in which customers, audiences, or users evaluate identifiable workers. External commentary may be informative, distorted, discriminatory, or unrelated to productive contribution, yet remain visible when managers make internal decisions. As personalized online evaluation becomes routine, understanding when digital social pressure is reflected in opportunity allocation and when organizations appropriately discount it remains important.

References

- Guy Aridor, Rafael Jiménez-Durán, Ro'ee Levy, and Lena Song. The economics of social media. *Journal of Economic Literature*, 62(4):1422–1474, 2024. doi: 10.1257/jel.20241743.
- Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social incentives in the workplace. *Review of Economic Studies*, 77(2):417–458, 2010.
- BBC Sport. Cristiano Ronaldo and Harry Maguire Most Abused Players on Twitter: Report. <https://www.bbc.co.uk/sport/football/62390281>, 2022. Accessed 30 July 2026.
- George Beknazar-Yuzbashev, Rafael Jiménez-Durán, Jesse McCrosky, and Mateusz Stalinski. Toxic content and user engagement on social media: Evidence from a field experiment. Working paper, May 2026.
- B. Douglas Bernheim. A theory of conformity. *Journal of Political Economy*, 102(5):841–877, 1994.
- Alasdair Brown, Dooruj Rambaccussing, J. James Reade, and Giambattista Rossi. Forecasting with social media: Evidence from tweets on soccer matches. *Economic Inquiry*, 56(3):1748–1763, 2018. doi: 10.1111/ecin.12506.
- Luis Cabral and Ali Hortacsu. The dynamics of seller reputation: Evidence from eBay. *Journal of Industrial Economics*, 58(1):54–78, 2010.
- Judith A. Chevalier and Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, 2006.
- Thomas Cornelissen, Christian Dustmann, and Uta Schönberg. Peer effects in the workplace. *American Economic Review*, 107(2):425–456, 2017.
- Thomas Davidson, Dana Warmusley, Michael Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM)*, pages 512–515. AAAI Press, 2017. doi: 10.1609/icwsml.v11i1.14955. URL <https://ojs.aaai.org/index.php/ICWSM/article/view/14955>.
- Stefano DellaVigna, John A. List, and Ulrike Malmendier. Testing for altruism and social pressure in charitable giving. *Quarterly Journal of Economics*, 127(1):1–56, 2012.

- Mathias Dewatripont, Ian Jewitt, and Jean Tirole. The economics of career concerns, part i: Comparing information structures. *The Review of Economic Studies*, 66(1):183–198, 1999.
- Luis Garicano, Ignacio Palacios-Huerta, and Canice Prendergast. Favoritism under social pressure. *Review of Economics and Statistics*, 87(2):208–216, 2005.
- Bengt Holmström. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):169–182, 1999. doi: 10.1111/1467-937X.00083.
- C. J. Hutto and Eric Gilbert. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 216–225. AAAI Press, 2014. doi: 10.1609/icwsm.v8i1.14550. URL <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>.
- IGN. “Lol Zero Spooked, Sorry”: Rian Johnson Responds to Kathleen Kennedy’s Suggestions That “Online Negativity” Put The Last Jedi Director Off Making More Star Wars. <https://www.ign.com/articles/lol-zero-spooked-sorry-rian-johnson-responds-to-kathleen-kennedys-suggestions-that-online-negativity-put-the-last-jedi-director-off-making-more-star-wars>, 2024. Accessed 30 July 2026.
- Ginger Zhe Jin and Phillip Leslie. Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics*, 1(1):237–267, 2009.
- Bing Liu. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012. doi: 10.2200/S00416ED1V01Y201204HLT016. URL <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>.
- Michael Luca. Reviews, reputation, and revenue: The case of Yelp.com. Technical Report 12-016, Harvard Business School NOM Unit, 2011.
- Michael Luca and Georgios Zervas. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, 62(12):3412–3427, 2016. doi: 10.1287/mnsc.2015.2304.
- Alexandre Mas and Enrico Moretti. Peers at work. *American Economic Review*, 99(1):112–145, 2009.
- Christine L. Porath and Amir Erez. Does rudeness really matter? the effects of rudeness on task performance and helpfulness. *Academy of Management Journal*, 50(5):1181–1197, 2007. doi: 10.5465/amj.2007.20159919.

Stefan Szymanski. A market test for discrimination in the English professional soccer leagues. *Journal of Political Economy*, 108(3):590–603, 2000. doi: 10.1086/262130.

The Guardian. Feeling the Force: Why Directing a Hollywood Blockbuster Is Tougher Than Ever. <https://www.theguardian.com/film/filmblog/2017/dec/15/blockbuster-star-wars-the-last-jedi-rian-johnson-feel-the-force>, 2017. Accessed 30 July 2026.

WhoScored.com. WhoScored Ratings Explained. <https://www.whoscored.com/explanations>, 2026. Accessed July 1, 2026.

Ellery Wulczyn, Nithum Thain, and Lucas Dixon. Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1391–1399. International World Wide Web Conferences Steering Committee, 2017. doi: 10.1145/3038912.3052591. URL <https://arxiv.org/abs/1610.08914>.

Table 1: Estimation-Sample Composition

Sample group	Appearance sample		Rating sample	
	Observations	Share	Observations	Share
Overall	38,653	100.0%	33,512	100.0%
<i>By league</i>				
Premier League	14,833	38.4%	12,934	38.6%
Other leagues	23,820	61.6%	20,578	61.4%
<i>By age</i>				
Age ≤ 26	20,411	52.8%	17,866	53.3%
Age > 26	18,242	47.2%	15,646	46.7%
<i>By nationality</i>				
Domestic	15,959	41.3%	13,991	41.7%
Foreign	22,694	58.7%	19,521	58.3%

Notes: The appearance sample contains Twitter-account player-match observations with a non-missing indicator for appearance in the team's next match within 14 days. The rating sample is smaller because it contains only observations for which the player makes a subsequent appearance within 14 days, allowing the rating in that appearance to be observed. Other leagues comprise the Bundesliga, La Liga, Ligue 1, and Serie A. Domestic status is defined relative to the league country. Shares are calculated within each outcome sample and split.

Table 2: Outcome Means by High versus Low Post-Match Online Negativity

Exposure group	Rating in next appearance Mean (SD)	Appearance probability Mean (SD)	Observations
<i>Panel A: Negative sentiment</i>			
Low exposure	6.910 (0.764)	0.902 (0.297)	28,975
High exposure	7.136 (0.892)	0.896 (0.305)	4,537
High minus low	0.226 [$p < 0.001$]	-0.006 [$p = 0.198$]	
<i>Panel B: Toxicity</i>			
Low exposure	6.918 (0.773)	0.901 (0.299)	28,749
High exposure	7.076 (0.853)	0.906 (0.292)	4,763
High minus low	0.158 [$p < 0.001$]	0.005 [$p = 0.246$]	
<i>Panel C: Racial abuse</i>			
Low exposure	6.931 (0.779)	0.901 (0.298)	31,896
High exposure	7.120 (0.899)	0.904 (0.295)	1,616
High minus low	0.189 [$p < 0.001$]	0.003 [$p = 0.705$]	

Notes: The table reports unconditional outcome means in the 33,512-observation main Twitter rating sample. The appearance probability is therefore calculated within the rating sample, not the full 38,653-observation appearance sample. For each online-negativity measure, high exposure is defined as a post-match level above that measure's sample mean. Low exposure is at or below the sample mean. Standard deviations are in parentheses. Brackets report p -values for tests of equal means. These descriptive comparisons are not causal because high- and low-exposure observations may differ systematically in player prominence, match context, prior form, and other characteristics.

Table 3: Online Negativity and Next-Team-Match Appearance: Player and Team-by-Current-Match Fixed Effects

	Single-measure specifications			Joint specification
	(1)	(2)	(3)	(4)
Panel A: Overall				
log(1 + negative sentiment mass)	-0.024*** (0.005)			-0.021*** (0.005)
log(1 + toxicity mass)		-0.020*** (0.006)		-0.013** (0.006)
log(1 + racial-abuse intensity)			-0.017* (0.010)	-0.002 (0.010)
Observations	38,653	38,653	38,653	38,653
Panel B: By age group				
<i>Age ≤ 26</i>				
log(1 + negative sentiment mass)	-0.024*** (0.006)			-0.020*** (0.007)
log(1 + toxicity mass)		-0.024*** (0.008)		-0.018** (0.008)
log(1 + racial-abuse intensity)			-0.017 (0.014)	-0.001 (0.014)
Observations	20,411	20,411	20,411	20,411
<i>Age > 26</i>				
log(1 + negative sentiment mass)	-0.025*** (0.007)			-0.024*** (0.008)
log(1 + toxicity mass)		-0.015* (0.008)		-0.009 (0.009)
log(1 + racial-abuse intensity)			-0.005 (0.014)	0.009 (0.014)
Observations	18,242	18,242	18,242	18,242
Player FE	✓	✓	✓	✓
Team-by-current-match FE	✓	✓	✓	✓
Rotation controls	✓	✓	✓	✓
Injury controls	✓	✓	✓	✓

Notes: Each column within a sample group is one regression. Panel A reports the overall sample. Panel B reports separate estimates for players aged 26 or younger and players older than 26. Columns (1)–(3) report single-measure specifications for negative sentiment, toxicity, and racial abuse, respectively. Column (4) is a joint specification that includes all three online-negativity measures. The three coefficient rows in column (4) within a sample group therefore come from the same regression. All specifications include player and team-by-current-match fixed effects, rotation controls, Transfermarkt injury controls, Twitter and pre-match exposure, prior form, next-match schedule characteristics, and player-specific controls not absorbed by fixed effects. Standard errors clustered by player are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4: Online Negativity and Next-Team-Match Appearance by League and Nationality Group

	Single-measure specifications			Joint specification
	(1)	(2)	(3)	(4)
Panel A: By league				
<i>Premier League</i>				
log(1 + negative sentiment mass)	-0.033*** (0.009)			-0.014 (0.011)
log(1 + toxicity mass)		-0.028*** (0.008)		-0.018* (0.010)
log(1 + racial-abuse intensity)			-0.024* (0.013)	-0.006 (0.012)
Observations	14,833	14,833	14,833	14,833
<i>Other leagues</i>				
log(1 + negative sentiment mass)	-0.018*** (0.006)			-0.014** (0.006)
log(1 + toxicity mass)		-0.033*** (0.011)		-0.026** (0.012)
log(1 + racial-abuse intensity)			-0.011 (0.016)	0.003 (0.017)
Observations	23,820	23,820	23,820	23,820
Panel B: By nationality				
<i>Domestic</i>				
log(1 + negative sentiment mass)	-0.029*** (0.007)			-0.030*** (0.007)
log(1 + toxicity mass)		-0.003 (0.010)		0.005 (0.010)
log(1 + racial-abuse intensity)			-0.003 (0.017)	0.005 (0.017)
Observations	15,959	15,959	15,959	15,959
<i>Foreign</i>				
log(1 + negative sentiment mass)	-0.020*** (0.007)			-0.012* (0.007)
log(1 + toxicity mass)		-0.032*** (0.007)		-0.025*** (0.007)
log(1 + racial-abuse intensity)			-0.031*** (0.012)	-0.012 (0.012)
Observations	22,694	22,694	22,694	22,694
Player FE	✓	✓	✓	✓
Team-by-current-match FE	✓	✓	✓	✓
Rotation controls	✓	✓	✓	✓
Injury controls	✓	✓	✓	✓

Notes: Each column within a sample group is one regression. Panel A reports separate estimates for the Premier League and the other four major European leagues. Panel B reports separate estimates for domestic and foreign players, where domestic means that the player's nationality code matches the league country. Columns (1)–(3) report single-measure specifications for negative sentiment, toxicity, and racial abuse, respectively. Column (4) is a joint specification that includes all three online-negativity measures. The three coefficient rows in column (4) within a sample group therefore come from the same regression. All specifications mirror Table 3 and include player and team-by-current-match fixed effects, rotation controls, Transfermarkt injury controls, Twitter and pre-match exposure, prior form, next-match schedule characteristics, and player-specific controls not absorbed by fixed effects. Standard errors clustered by player are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Online Negativity and the Rating in the Player's Next Appearance: Overall Specifications and Age Groups

	Single-measure specifications			Joint specification
	(1)	(2)	(3)	(4)
Panel A: Overall, player and team-by-current-match fixed effects				
log(1 + negative sentiment mass)	-0.015* (0.009)			-0.013 (0.009)
log(1 + toxicity mass)		-0.011 (0.011)		-0.007 (0.012)
log(1 + racial-abuse intensity)			-0.005 (0.020)	0.004 (0.021)
Observations	33,512	33,512	33,512	33,512
Player FE	✓	✓	✓	✓
Team-by-current-match FE	✓	✓	✓	✓
Rotation controls	✓	✓	✓	✓
Injury controls	✓	✓	✓	✓
Panel B: Overall, team-by-current-match fixed effects only				
log(1 + negative sentiment mass)	0.009 (0.008)			0.007 (0.009)
log(1 + toxicity mass)		0.010 (0.010)		0.008 (0.011)
log(1 + racial-abuse intensity)			0.010 (0.021)	0.002 (0.023)
Observations	33,512	33,512	33,512	33,512
Player FE	×	×	×	×
Team-by-current-match FE	✓	✓	✓	✓
Rotation controls	✓	✓	✓	✓
Injury controls	✓	✓	✓	✓
Panel C: By age group, player and team-by-current-match fixed effects				
<i>Age ≤ 26</i>				
log(1 + negative sentiment mass)	-0.007 (0.013)			-0.009 (0.013)
log(1 + toxicity mass)		0.007 (0.016)		0.015 (0.017)
log(1 + racial-abuse intensity)			-0.022 (0.028)	-0.028 (0.028)
Observations	17,866	17,866	17,866	17,866
<i>Age > 26</i>				
log(1 + negative sentiment mass)	-0.036*** (0.013)			-0.037*** (0.013)
log(1 + toxicity mass)		-0.020 (0.016)		-0.018 (0.018)
log(1 + racial-abuse intensity)			0.043 (0.028)	0.066** (0.030)
Observations	15,646	15,646	15,646	15,646
Player FE	✓	✓	✓	✓
Team-by-current-match FE	✓	✓	✓	✓
Rotation controls	✓	✓	✓	✓
Injury controls	✓	✓	✓	✓

Notes: Each column within a sample group is one regression. Panels A and B report the overall sample. Panel A includes player and team-by-current-match fixed effects. Panel B omits player fixed effects and retains team-by-current-match fixed effects. Panel C reports separate estimates for players aged 26 or younger and players older than 26, with player and team-by-current-match fixed effects. Columns (1)–(3) report single-measure specifications for negative sentiment, toxicity, and racial abuse, respectively. Column (4) is a joint specification that includes all three online-negativity measures. The three coefficient rows in column (4) within a sample group therefore come from the same regression. All specifications mirror the appearance specifications in Tables 3 and 4, replacing the appearance outcome with the rating in the player's next appearance. Standard errors clustered by player are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 6: Online Negativity and the Rating in the Player's Next Appearance by League and Nationality Group

	Single-measure specifications			Joint specification
	(1)	(2)	(3)	(4)
Panel A: By league				
<i>Premier League</i>				
log(1 + negative sentiment mass)	-0.021 (0.018)			-0.055** (0.026)
log(1 + toxicity mass)		0.003 (0.013)		0.035* (0.020)
log(1 + racial-abuse intensity)			0.002 (0.026)	0.007 (0.027)
Observations	12,934	12,934	12,934	12,934
<i>Other leagues</i>				
log(1 + negative sentiment mass)	-0.016 (0.011)			-0.010 (0.011)
log(1 + toxicity mass)		-0.046** (0.022)		-0.044* (0.024)
log(1 + racial-abuse intensity)			0.003 (0.031)	0.022 (0.032)
Observations	20,578	20,578	20,578	20,578
Panel B: By nationality				
<i>Domestic</i>				
log(1 + negative sentiment mass)	-0.020 (0.016)			-0.020 (0.016)
log(1 + toxicity mass)		-0.014 (0.021)		-0.014 (0.021)
log(1 + racial-abuse intensity)			0.022 (0.030)	0.035 (0.029)
Observations	13,991	13,991	13,991	13,991
<i>Foreign</i>				
log(1 + negative sentiment mass)	-0.020 (0.013)			-0.020 (0.014)
log(1 + toxicity mass)		-0.007 (0.013)		-0.002 (0.015)
log(1 + racial-abuse intensity)			0.006 (0.028)	0.014 (0.030)
Observations	19,521	19,521	19,521	19,521
Player FE	✓	✓	✓	✓
Team-by-current-match FE	✓	✓	✓	✓
Rotation controls	✓	✓	✓	✓
Injury controls	✓	✓	✓	✓

Notes: Each column within a sample group is one regression. Panel A reports separate estimates for the Premier League and the other four major European leagues. Panel B reports separate estimates for domestic and foreign players, where domestic means that the player's nationality code matches the league country. Columns (1)–(3) report single-measure specifications for negative sentiment, toxicity, and racial abuse, respectively. Column (4) is a joint specification that includes all three online-negativity measures. The three coefficient rows in column (4) within a sample group therefore come from the same regression. All specifications mirror Table 5 and include player and team-by-current-match fixed effects, rotation controls, Transfermarkt injury controls, Twitter and pre-match exposure, prior form, next-match schedule characteristics, and player-specific controls not absorbed by fixed effects. Standard errors clustered by player are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7: Winner–Loser Differences in Changes in Online Negativity: Close-Match Local Linear Comparisons

	Negative sentiment change		Toxicity change		Racial-abuse change	
	BW 0.25 (1)	BW 0.50 (2)	BW 0.25 (3)	BW 0.50 (4)	BW 0.25 (5)	BW 0.50 (6)
Unweighted	-0.0166*** (0.0027)	-0.0201*** (0.0018)	-0.0020*** (0.0002)	-0.0020*** (0.0002)	-0.0001 (0.0001)	0.0001 (0.0002)
Total-tweet weighted	-0.0200*** (0.0021)	-0.0208*** (0.0017)	-0.0021*** (0.0003)	-0.0024*** (0.0002)	-0.0003 (0.0002)	-0.0001* (0.0001)
Post-tweet weighted	-0.0209*** (0.0022)	-0.0214*** (0.0019)	-0.0024*** (0.0003)	-0.0027*** (0.0002)	-0.0003 (0.0002)	-0.0001 (0.0001)
Observations	7,327	16,278	7,327	16,278	7,327	16,278
Player FE	✓	✓	✓	✓	✓	✓
Match FE	✓	✓	✓	✓	✓	✓

Notes: The table reports conditional winner–loser differences in the change in online negativity from the 24 hours before kickoff to the 24 hours after the match. Samples are restricted to expected-goal differences, xGD , within bandwidths of 0.25 or 0.50 around zero. Each cell reports the win coefficient from a separate local linear regression with a linear xGD term, a win-by- xGD interaction, player fixed effects, match fixed effects, and the controls in Equation (3). Total-tweet weights count tagged tweets from hours –24 through 24. Post-tweet weights count tagged tweets from hours 0 through 24. Both weights are floored at one. Unweighted regressions assign weight one to each player-match. Standard errors clustered at the match level are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

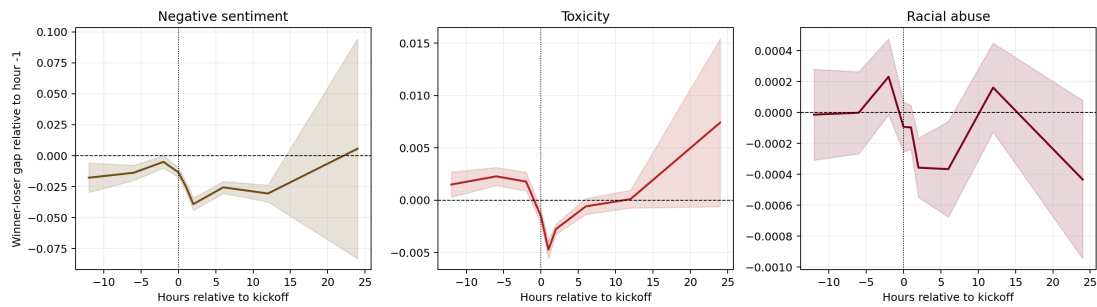


Figure 1: Winner–Loser Differences in Online Negativity by Event Hour

Notes: The figure plots selected-hour coefficients on the interactions between winning and event hour, with hour –1 omitted. Specifications include player, match, and event-hour fixed effects. The sample contains hourly observations whose team-match identifiers appear in the main Mahalanobis no-replacement matched sample. Appendix Table A5 reports the coefficient estimates.

Table A1: Transfermarkt Injury Share by League and Season in the Appearance Sample

League	Season	Observed player-dates	Injured player-dates	Share injured
Bundesliga	2017–2018	1,515	71	0.047
Bundesliga	2018–2019	1,431	76	0.053
Bundesliga	2019–2020	1,390	61	0.044
Bundesliga	2020–2021	1,321	67	0.051
La Liga	2017–2018	2,222	87	0.039
La Liga	2018–2019	2,462	104	0.042
La Liga	2019–2020	2,210	75	0.034
La Liga	2020–2021	2,371	87	0.037
Ligue 1	2017–2018	903	22	0.024
Ligue 1	2018–2019	879	30	0.034
Ligue 1	2019–2020	567	28	0.049
Ligue 1	2020–2021	1,018	37	0.036
Premier League	2017–2018	3,395	87	0.026
Premier League	2018–2019	3,496	151	0.043
Premier League	2019–2020	2,831	123	0.043
Premier League	2020–2021	3,777	140	0.037
Serie A	2017–2018	970	21	0.022
Serie A	2018–2019	1,014	32	0.032
Serie A	2019–2020	942	35	0.037
Serie A	2020–2021	1,393	72	0.052

Notes: The table reports Transfermarkt-observed player-dates in the main Twitter-only appearance sample by league and season. Injured player-dates are next-team-match player-dates classified as injuries by the Transfermarkt roster-status field used to construct the injury controls.

Table A2: Summary Statistics for the Analysis Sample

Statistic	Level	Value
<i>Panel A: Overall counts</i>		
Footballers in linked panel	Player	3,930
Footballers with Twitter account	Player	652
Share with Twitter account	Player	16.6%
Linked player-match observations	Player-match	160,194
Twitter player-match observations	Player-match	42,303
Main Twitter rating outcome sample	Player-match	33,988
Main rating estimation sample	Player-match	33,512
Main appearance estimation sample	Player-match	38,653
<i>Panel B: Player characteristics</i>		
Mean age	Player-match	25.96
SD age	Player-match	4.02
Players with non-missing skin tone	Player	3,468
Skin-tone coverage	Player	88.2%
Mean skin-tone score	Player	0.465
SD skin-tone score	Player	0.107
Share skin tone > 0.5	Player	31.2%
<i>Panel C: League composition</i>		
Premier League	Player-league	887
Serie A	Player-league	923
Bundesliga	Player-league	804
La Liga	Player-league	862
Ligue 1	Player-league	968
<i>Panel D: Nationality composition</i>		
Domestic players	Player-league	1,869
Foreign players	Player-league	2,231
Share foreign	Player-league	57.9%

Notes: The table summarizes the linked five-league analysis panel and the main Twitter-account estimation samples. League rows count player-league appearances, so a player appearing in two leagues can contribute to two league rows. Nationality composition uses the domestic-versus-foreign rule applied in the regressions.

Table A3: Positive Sentiment, Online Negativity, and Subsequent Player Outcomes

	Next-team-match appearance		Rating in next appearance	
	Single positive	Four-measure joint	Single positive	Four-measure joint
	(1)	(2)	(3)	(4)
log(1 + negative sentiment mass)		-0.024*** (0.005)		-0.014 (0.009)
log(1 + positive sentiment mass)	0.006 (0.005)	0.017*** (0.005)	-0.002 (0.009)	0.003 (0.009)
log(1 + toxicity mass)		-0.017*** (0.006)		-0.008 (0.012)
log(1 + racial-abuse intensity)		-0.005 (0.010)		0.003 (0.021)
Observations	38,653	38,653	33,512	33,512
Player FE	✓	✓	✓	✓
Team-by-current-match FE	✓	✓	✓	✓
Rotation controls	✓	✓	✓	✓
Injury controls	✓	✓	✓	✓

Notes: Columns (1) and (3) report single-measure specifications for positive sentiment. Columns (2) and (4) report four-measure joint specifications that include negative sentiment, positive sentiment, toxicity, and racial abuse in the same regression. Positive sentiment mass sums the positive VADER sentiment component over the same tagged post-match tweets used to construct negative sentiment mass. All four measures enter as log(1 + mass) or log(1 + intensity). The specifications mirror Panel A of Tables 3 and 5, with player and team-by-current-match fixed effects, rotation controls, Transfermarkt injury controls, Twitter and pre-match exposure, prior form, next-match schedule characteristics, and player-specific controls not absorbed by fixed effects. Standard errors clustered by player are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

**Table A4: Winner–Loser Differences in Changes in Online Negativity:
Matched-Sample Robustness**

Matched sample	Negative sentiment change (1)	Toxicity change (2)	Racial-abuse change (3)
Mahalanobis, no replacement	-0.0165*** (0.0025)	-0.0009*** (0.0002)	-0.0004* (0.0002)
Mahalanobis, with replacement	-0.0173*** (0.0011)	-0.0035*** (0.0002)	-0.0001 (0.0001)
Propensity score, no replacement	-0.0167*** (0.0035)	-0.0012*** (0.0003)	0.0003 (0.0003)
Propensity score, with replacement	-0.0183*** (0.0017)	-0.0022*** (0.0002)	-0.0001 (0.0001)
Pair FE	✓	✓	✓
Player FE	✓	✓	✓
Match FE	✓	✓	✓

Notes: The table reports conditional winner–loser differences in the change in online negativity from the 24 hours before kickoff to the 24 hours after the match. Each cell reports the win coefficient from a separate weighted regression that absorbs matched-pair, player, and match fixed effects. The Mahalanobis no-replacement, Mahalanobis with-replacement, propensity-score no-replacement, and propensity-score with-replacement samples contain 6,752 pairs and 13,504 observations, 15,260 pairs and 30,520 observations, 6,794 pairs and 13,588 observations, and 15,241 pairs and 30,482 observations, respectively. Standard errors clustered by player are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A5: Winner–Loser Differences in Online Negativity by Event Hour

Hour	Negative sentiment	Toxicity	Racial abuse
–12	–0.0178*** (0.0060)	0.0015** (0.0006)	–0.0000 (0.0002)
–6	–0.0138*** (0.0031)	0.0023*** (0.0004)	–0.0000 (0.0001)
–2	–0.0049** (0.0025)	0.0018*** (0.0005)	0.0002* (0.0001)
0	–0.0135*** (0.0020)	–0.0015*** (0.0003)	–0.0001 (0.0001)
1	–0.0248*** (0.0018)	–0.0047*** (0.0004)	–0.0001 (0.0001)
2	–0.0391*** (0.0026)	–0.0028*** (0.0003)	–0.0004*** (0.0001)
6	–0.0255*** (0.0025)	–0.0006 (0.0004)	–0.0004** (0.0002)
12	–0.0305*** (0.0035)	0.0001 (0.0004)	0.0002 (0.0001)
24	0.0057 (0.0453)	0.0074* (0.0041)	–0.0004* (0.0003)
Observations	93,971	93,971	93,971
Player FE	✓	✓	✓
Match FE	✓	✓	✓
Event-hour FE	✓	✓	✓

Notes: The table reports selected-hour winner–loser differences in online negativity. Each cell reports the coefficient on the interaction between the win indicator and the listed event hour, with hour –1 omitted. The sample contains hourly player-match observations whose team-match identifiers appear in the main Mahalanobis no-replacement matched sample. Specifications include player, match, and event-hour fixed effects. Standard errors clustered by player are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.