Exploring entertainment utility from football games

Article

Published Version

Pawlowski, T., Rambacussing, D., Ramirez, P., Reade, J. J. ORCID: https://orcid.org/0000-0002-8610-530X and Rossi, G. (2024) Exploring entertainment utility from football games. Journal of Economic Behaviour & Organisation, 223. pp. 185-198. ISSN 0167-2681 Available at https://centaur.reading.ac.uk/116122/

Publisher: Elsevier

Journal of Economic Behavior and Organization 223 (2024) 185–198

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Research paper

Exploring entertainment utility from football games☆

Tim Pawlowski [a,1], Dooruj Rambaccussing [b], Philip Ramirez [c], J. James Reade [c,*], Giambattista Rossi [d]

[a] *University of Tübingen, Wilhelmstraße 124, DE 72074, Tübingen, Germany*
[b] *University of Dundee, 1–3 Perth Road, United Kingdom*
[c] *University of Reading, Whiteknights, PO Box 217, United Kingdom*
[d] *Birkbeck University of London, Malet Street, Bloomsbury, United Kingdom*

ARTICLE INFO

Dataset link: https://github.com/philiprami/surprise_suspense_sentiment_shock

JEL classification:
C10
D91
L83

Keywords:
Suspense
Surprise
Entertainment utility
football
Tweets

ABSTRACT

Previous research exploring the role of belief dynamics for consumers in the entertainment industry has largely ignored the fact that emotional reactions are a function of the content *and* a consumer's disposition towards certain protagonists. By analyzing 19 m tweets in combination with in-play information for 380 football matches played in the English Premier League we contribute to the literature in three ways. First, we present a setting for testing how belief dynamics drive behavior which is characterized by several desirable features for empirical research. Second, we present an approach for detecting *fans* and *haters* of a club as well as *neutrals* via sentiment revealed in Tweets. Third, by looking at behavioral responses to the temporal resolution of uncertainty during a game, we offer a fine-grained empirical test for the popular uncertainty-of-outcome hypothesis in sports.

1. Introduction

Modeling dynamic choice problems with an explicit focus on uncertainty attached to a certain point in time goes back to Kreps and Porteus (1978), who explored preferences for the earlier or later resolution of uncertainty. Several scholars have since extended these ideas. For instance, Palacios-Huerta (1999) has focused on the form of the timing of the resolution by explicitly modeling disappointment aversion, as introduced by Gul (1991). This model can explain a preference for the one-shot rather than the sequential resolution of uncertainty (for further extensions, see Kőszegi and Rabin, 2009, or Dillenberger, 2010). Caplin and Leahy (2001) more broadly considered both negative and positive anticipatory emotions felt by individuals before uncertainty is resolved. For instance, they define *suspense* as a positive anticipatory emotion which might explain why fans in sports bet on their favorite team as observed by Babad and Katz (1991), i.e., fans simply want to increase their feelings of suspense.

This literature informed the seminal work by Ely et al. (2015) who modeled the demand for non-instrumental information by focusing on entertainment utility from *suspense* and *surprise*. While suspense is attributed to the variance in the next period's beliefs, thus representing a forward-looking measure, surprise results from an outcome that contradicts anterior beliefs representing

☆ Dooruj Rambaccussing, James Reade, and Giambattista Rossi acknowledge support from Twitter through their #DataGrants scheme.

* Corresponding author.
E-mail addresses: tim.pawlowski@uni-tuebingen.de (T. Pawlowski), d.rambaccussing@dundee.ac.uk (D. Rambaccussing), p.ramirez@pgr.reading.ac.uk (P. Ramirez), j.j.reade@reading.ac.uk (J.J. Reade), g.rossi@bbk.ac.uk (G. Rossi).
[1] Tim Pawlowski is also affiliated with the LEAD Graduate School and Research Network as well as the Interfaculty Research Institute for Sports and Physical Activity in Tuebingen.

https://doi.org/10.1016/j.jebo.2024.04.018
Received 7 July 2023; Received in revised form 11 February 2024; Accepted 17 April 2024
Available online 23 May 2024

a backward-looking measure. The authors close by writing: *"How suspense, surprise, and other aspects of belief dynamics drive demand for noninstrumental information is fundamentally an empirical question, one that we hope will be addressed by future research"* (Ely et al., 2015).

Only a small number of researchers to date have followed their call by empirically exploring this in sports. Bizzozero et al. (2016) used minute-by-minute TV viewing figures from 80 Wimbledon men's singles tennis matches and operationalized *suspense* and *surprise* with information coming from betting markets. Buraimo et al. (2020) used minute-by-minute TV viewing figures for 540 Premier League matches and added a further concept, *shock*. Instead of relying on in-play odds from betting markets, they derived implied probabilities for each outcome in each minute by feeding an in-play model. Richardson et al. (2023) replicated this study using minute-by-minute TV viewing figures for 180 (131) UEFA Champions League games televised in the UK (Spanish) market. Kaplan (2021) used 15-minute interval TV ratings from 477 National Basketball Association (NBA) games during the 2017–18 and 2018–19 seasons and compared the impact of *thrilll* (measured by *suspense* and *surprise*) and *skilll* (measured by productivity and popularity). Simonov et al. (2022) used detailed viewership information for a sample of 104 professional eSport tournament games summing up to more than 2,700 rounds played. These data allow modeling the decision-to-join and the decision-to-leave a (Twitch.tv) stream separately. Finally, Liu et al. (2021) used individual-level data about 877 baseball telecasts during the 2018 Japanese Major League season. The granular data which were built, amongst others, upon utilizing a facial recognition algorithm, allow to further disentangle the effects of *suspense* and *surprise* for actively versus passively attentive viewers. In fact, many consumers often do not pay full attention to the television programming since, for instance, they might actively search for game-related information and/or just do different things in parallel, such as cooking, or tweeting about the game.

While these studies find that *suspense* and (to some extent) also *surprise* and *shock* are important drivers of demand, detailed findings reveal some interesting and partly contradictory issues. For instance, (i) while Bizzozero et al. (2016) find that *surprise* has a larger impact than *suspense* in tennis, Kaplan (2021), Buraimo et al. (2020) and Richardson et al. (2023) as well as Simonov et al. (2022) find the opposite pattern in basketball, football and eSports respectively. (ii) *Suspense* decreases the probability of leaving a stream while neither *surprise* nor *suspense* unfold any effects on the decision to join a stream (Simonov et al., 2022). (iii) *Suspense* and *surprise* seem to primarily impact viewership on the intensive margin, i.e., within games (Kaplan, 2021). (iv) Spectators have a higher probability to turn on games featuring less popular players/teams if they are nearing the end *and* exhibiting sufficiently high *suspense* (Kaplan, 2021). Finally, (v) postseason games amplify the effects of *suspense* and *surprise* (Kaplan, 2021), and (vi) women seem to be less responsive to *suspense* and *surprise* than men (Liu et al., 2021).

Despite the contribution of these studies to better understand how entertainment utility translates into the demand for sports, two main shortcomings exist which we intend to address in this study. First, the setting analyzed, i.e., TV/stream viewing behavior, requires a careful distinction between the decision-to-join versus the decision-to leave a program/stream (Simonov et al., 2022) and between active versus passive viewing (Liu et al., 2021). While some studies try to approach these issues with more fine-grained data and complex measures, we propose analyzing a more simple setting: social media behavior, and in particular behavior on Twitter,[2] where individuals decide whether to post a Tweet.[3] Second, neither of the studies is able to reveal whether and how fandom is moderating the relation of interest since TV/stream viewing figures do not allow any distinction between fans. However, according to affective disposition theory (Zillmann and Cantor, 1972), emotional reactions by fans are a function of the content *and* a fan's disposition towards athletes/teams in contention (Raney, 2018).

We approach both shortcomings by combining data on in-game events with betting odds and Tweets for 380 games played in the English Premier League (EPL) in season 2013/2014. While the former two data sets are used for operationalizing *surprise*, *suspense*, and *shock*, the latter data allow us to derive temporal sentiment and distinguish between different types of individuals. We start with generating sentiment scores for each Tweet using comprehensive algorithm-based approaches by employing Natural Language Processing (NLP). The calculated average post-game sentiment scores for every Twitter user enable us to identify *fans* and *haters* of a club as well as *neutrals* for each game. In order to explore entertainment utility from football games for these different types of individuals, we regress the number of Tweets per minute on *surprise*, *suspense*, and *shock*. Moreover, we explore asymmetries in behavior by disentangling the effects for *fans* and *haters* when 'their' team is losing or winning.

Our findings suggest that emotional cues significantly influence the number of Tweets in a given minute. While both *surprise* and *shock increase* the number of Tweets, *suspense decreases* the number of Tweets which could be explained by individuals being 'caught in the moment' probably leaving no time to tweet. Moreover, as could be expected, any response to emotional cues is smallest for *neutrals*. Further analysis reveals that goal-induced *immediate* effects from *surprise* and *shock* on Twitter activity are the largest, when the favorite (or hated) team either scores or concedes an equalizer. A closer look into these findings reveals some interesting similarities in behavior between *fans* and *haters*. Goal-induced *immediate* effects from *surprise* are particularly small when the probability of the 'desired' result — i.e., winning of the favorite team (for *fans*) and losing of the hated team (for *haters*) — increases. In principle, this could be suggestive of some sort of extraordinary celebration taking place in the minute of the 'desired' goal, thus eventually postponing Twitter activity.[4] More importantly, however, this observation adds further credibility to

[2] Since we make use of historic data we keep Twitter instead of X as the platform name in our paper.

[3] Note that exploring the effects of emotional cues on Twitter activity was already proposed by Kaplan (2021) who writes on p. 16: *"Future work can directly assess the relevance of each of these mechanisms using household-level viewership data as well as complementary data from information-providing applications (e.g. Twitter)"*. Yet, the only study investigating the effects of emotional cues on complementary activities beyond watching is Fischer et al. (2023). They explored the effects of *suspense* and *surprise* on alcohol consumption during a match.

[4] Note, that we refrain from further exploring any lagged effects in our setting given econometric concerns caused by the temporal structure of the data with many measurement points.

our identification exercise because one would expect that both *fans* and *haters* respond in a similar way when the probability of winning of the favorite team (for *fans*) and losing of the hated team (for *haters*) increases.

Overall, these findings could inform the literature in three ways. First, we present a novel setting for testing whether and how belief dynamics drive behavior. This seems highly relevant given the lack of research about immediate emotions and the consequences of a wide range of visceral factors for (immediate) human behavior *in general* (Loewenstein, 2000). Moreover, this seems promising given the identified drawbacks when modeling, for instance, TV/stream viewing behavior as discussed before. Second, we present an approach for detecting *fans* and *haters* of a club as well as *neutrals* via sentiment revealed in Tweets. From a managerial perspective this approach might help to further develop and implement personalized forms of communication by clubs and sponsors.[5] Third, by looking at behavioral responses to the temporal resolution of uncertainty during the course of a game, we offer a different and more fine-grained type of empirical test for the well-known uncertainty-of-outcome hypothesis in sports.[6] This seems relevant from a policy perspective, since the hypothesis still lacks empirical support even though it forms the basic argument for all cross-subsidization measures and labor market interventions in professional sport leagues around the globe (see, for instance, Pawlowski et al. (2018)).[7] Our findings suggest that entertainment utility is influenced by elements which gain in (lagged) *certainty* (such as surprise or shock) as well as elements which gain in *uncertainty* (such as suspense). In this regard, the negative effect of *suspense* is suggestive of individuals paying more attention to the match itself instead of complementary activities like Tweeting. This proposition is fully backed up by studies exploring the demand for sports telecasts which unambiguously reveal a positive effect of *suspense* on viewing figures. Moreover, it is in line with Fischer et al. (2023) who find that *suspense* reduces alcohol purchases in the stadium during a match.

2. Data and method

2.1. Identifying fans, haters, and neutrals

Our data comprise of all worldwide English-language Tweets that mention any hashtags associated with a team in the EPL before, during, and after all 380 matches played in the 2013/2014 season. More precisely, Tweets in our data are associated with a given match because of the temporal overlap with match day/time plus the use of hashtags for either team in contention.[8] This amounts to about 19 million unique Tweets for our analysis.

Our objective is distinguishing Tweets, that were posted either by a *fan* of the home team, a *fan* of the away team, a *hater* of the home team, a *hater* of the away team, or a *neutral*. However, our raw data only allows distinguishing Tweets using a home team hashtag and those using an away team hashtag (or both). While it seems reasonable to assume that a *fan* of a team commonly uses a hashtag of the team she supports, i.e., for instance, a home team *fan* uses a home team hashtag, we must keep in mind that a home *fan* might also use a hashtag of the away team or hashtags of both teams. Moreover, also *haters* of the home team/away team as well as *neutrals* either use a home team and/or an away team hashtag. In other words, we have (home/away team) *fans*, (home/away team) *haters*, and *neutrals* posting Tweets with a home team hashtag as well as (home/away team) *fans*, (home/away team) *haters*, and *neutrals* posting Tweets with an away team hashtag and those posting Tweets with multiple hashtags. This means that just distinguishing between home and away team hashtags in our analysis does not allow for discriminating between the different types of *fans* and *haters*, or *neutrals*. In this study, we propose exploiting the sentiment expressed in Tweets in order to identify the different types of Twitter users. This is further explained below.

As a first step, we need to measure sentiment. While a range of ways of measuring sentiment exist – from simply assigning words a positive or negative number, to classifying particular passages of words as being positive or negative – we make use of comprehensive algorithm-based approaches by employing Natural Language Processing (NLP). Overall, we explore and compare two distinct NLP-approaches. The first approach generates sentiment scores, ranging from 0 (very negative) to 25 (very positive), for each Tweet using a Random Forest (RF) estimator trained on data from the Stanford's Sentiment Tree Bank. Broadly speaking, the RF estimator produces an ensemble of decision trees popularly used for NLP. The main advantage of this self-trained estimator is that we can monitor how important individual features are in determining outcomes. Our model was trained on more than three million features or word tokens with the ten most important features being *bad*, *performance*, *best*, *n't*, *funny*, *dull*, *great*, *like*, *good*, and *waste* (see Appendix A.4 for further details on the architecture of the winning model).

The second approach builds upon a pre-trained high-performing algorithm, i.e., the 'Bidirectional Encoder Representations from Transformers' (BERT) model (Devlin et al., 2018). Developed by researchers at Google AI Language in 2018, this deep learning model

[5] For a recent discussion on the personal, social, and commercial relevance of understanding such behavior, see Jiwa et al. (2021).

[6] The uncertainty-of-outcome hypothesis (UOH) originates from the seminal works by Rottenberg (1956) and Neale (1964) and suggests a positive relation between the level of uncertainty over the outcome of a sports competition and its attractiveness for spectators and fans.

[7] To the best of our knowledge, only one study exists that has used Twitter data for testing the UOH before. Lucas et al. (2017) use three different types of information about 60 FIFA World Cup games in 2014, i.e., Vegas betting odds in order to measure differences between predicted and actual scores for both teams, a game's average Tweets per minute as a proxy for attendance by/excitement of the Twitter audience, and the proportion of Tweets which were positive, negative, or neutral during a game. Simple game-level correlations reveal, that games with bigger than expected score differences had higher Tweets per minute and a higher share of negative Tweets. They argue, that the latter finding is in line with the UOH while the former contradicts the UOH. We argue, however, that game-level correlations do not provide any credible evidence. Moreover, the authors did not make use of the elaborated cue measures as proposed by Ely et al. (2015) and partly even confuse emotional cue and attention measures.

[8] Taking the example of Liverpool, a corresponding Tweet contains one or more of FC Liverpool, @LFC, @lfcbuzztap, @empireofthekop, @liverpool, @Liverpool_FC_, @thisisanfield, #lfc, #liverpool, #liverpoolfc, or #ynwa. For further details about data and methods, please see Appendix A.

serves as a powerful solution for many common NLP tasks such as sentiment analysis and named entity recognition. In contrast to previous machine learning models analyzing text sequence using left-to-right (or right-to-left) training, the bidirectionally trained language model demonstrates a much deeper understanding of language.[9] In contrast to the RF model, the BERT model classifies each Tweet into three possible categories, i.e., positive, negative, and neutral. Alongside each prediction, the transformer returns the corresponding probability of the predicted label. Since we are largely concerned with positive sentiment post-*win*/post-*loss* (as will be explained next), the probability scores for positive Tweets translate well to our predefined rules to determine *fans* and *haters*, making both approaches comparable.

In the next step, we isolate post-*win* and post-*loss* sentiment scores for every Twitter user for each game. For identifying *fans*, we rank the average post-*win* sentiment scores per game and take the most commonly occurring team in a user's Top-5.[10] If the user does not comment positively on more than one win of a particular team, we fail to assign fandom for this user. In other words, a team must appear at least twice in a user's Top-5 in order to be considered. Conversely, to determine a *hater*, we exploit post-*loss* sentiment expressed in Tweets. Again, ranked by average sentiment score of tweets, we take the most commonly occurring team in each user's Top-5. The intuition behind this approach is that positive sentiment after a loss probably reflects some kind of *schadenfreude*. Like for fandom, if the Twitter user does not take delight in at least two losses for a team, hatership cannot be found. The remainder of users are neither denoted as a *fan* nor as a *hater* and are assumed to be *neutral*.

Because most users tweet about a team post-*win* rather than post-*loss* – with eligible users (users with at least three Tweets) tweeting 1,145,281 times about the winner of a team post-*win* and only 598,799 times about the loser of a match post-*loss* – the user's top scoring is heavily favored in determining fandom vs. hatership. Moreover, post-*win*, the average top scoring sentiment is 14.59, well above the average sentiment score overall of 13.51. Post-*loss*, however, the average top scoring sentiment is 13.77 which is just above average. This suggests, more often than not, a user delights on their own team's success more than celebrates another's demise, thus, making it generally easier to assign fandom as opposed to hatership. In principle, this procedure allows assigning fandom and hatership of two *different* teams for the *same* Twitter user. If, however, we find overlap between fandom and hatership of the *same* team for a Twitter user (many users regularly comment on just one or two clubs), we assign either fandom or hatership according to the higher absolute value of the post-match sentiment score. As such, if the user has a higher average sentiment score post-*win*, the user is determined to be a *fan*. If the higher average sentiment score occurs post-*loss*, the user is marked as a *hater*.

Following this procedure, we identified amongst all 1.3 million Twitter users 200,062 *fans* and 27,539 *haters* according to the RF model (hereof, only 6,281 users were identified as both a *fan* of one team and a *hater* of another team). Fig. 1 provides a breakdown of these numbers for each team. The remaining 1,096,225 users are regarded as *neutrals*.[11] Since the BERT model is able to detect more nuance than the RF model in the data, it is not surprising that the BERT model predicts less *fans* (135,777) than the RF model. However, it is notable that more than 80% of the predicted *fans* by the BERT model are the same as predicted by the RF model. Moreover, only about 0.5% were predicted by the BERT model as a *fan* of another team. These tendencies are similar but more pronounced regarding *haters*. Overall, the BERT model predicts many fewer *haters* (7,132) than the RF model. However, roughly half of the *haters* predicted by the BERT model are the same as predicted by the RF model and less than 0.1% were predicted by the BERT model as a *hater* of another team. Since we have no strong arguments in favor or against either model, we proceed with our analysis by distinguishing between *fans/haters* as defined by the RF model (200,062/27,539), the BERT model (135,777/7,132), as well as those predicted by both models, i.e., the overlap samples (108,759/3,534).

In order to get a first impression of how well this classification exercise works, we take an example from the match between Liverpool and Chelsea on matchday 36 of 38. The match was critical for the championship race and ended with a 0–2 home loss leaving Liverpool with a considerably reduced chance of winning the title. Out of overall 206,411 'Liverpool' Tweets before, during, and after this match, approximately every fourth was posted by a Liverpool *fan* as identified by our approach. However, as expected, of the 5,198 users retweeting "@LFC LOL!" after the game, only a marginal portion of these users are Liverpool *fans* as identified by our approach.[12]

More generally, we find some strong correlations between the overall number of fans identified by our approach and the average number of spectators attending matches of each team (see Fig. 2) as well as the number of (actual) followers of the official team accounts (see Fig. 3) adding some further credibility to our approach.[13]

[9] Specifically, we deploy a transformer featured in the huggingface library named bertweet-base-sentiment-analysis. Developed by pysentimiento, the model was pre-trained with SemEval 2017 corpus (around 40,000 tweets).

[10] While choosing the Top-5 may initially appear like an arbitrary cutoff, any variation in this cutoff number has little impact to our outcomes. This is because the vast majority (95%) of Twitter users in our dataset tweet about three different teams or less (98% of Twitter users comment on five or less teams). In fact, if we change this cutoff from five to ten, there is only a difference in *fan* prediction of 0.5%. A switch to the Top-3 demonstrates an even smaller impact, i.e., 0.3%.

[11] The corresponding Figures according to the transformer model (BERT) and the overlap sample identified by both models can be found in Appendix B Figs. B1 and B2.

[12] More precisely, 27% (23%) [19%] of overall 206,411 'Liverpool' Tweets are by Liverpool *fans* while only 2.14% (1.65%) [1.12%] of the 5,198 users retweeting "@LFC LOL!" after the game are Liverpool *fans* according to the RF model (the BERT model) [the overlap sample from both models].

[13] Note that the counts of followers were taken in March 2022, i.e., several years after the Tweets.

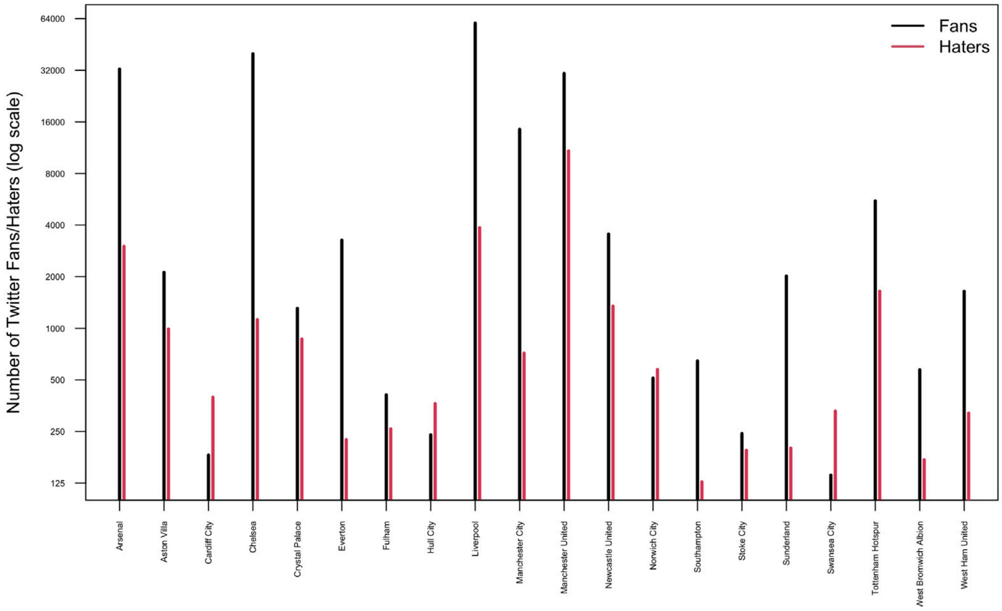

Fig. 1. Twitter users regarded as *fans* and *haters* according to the RF model.

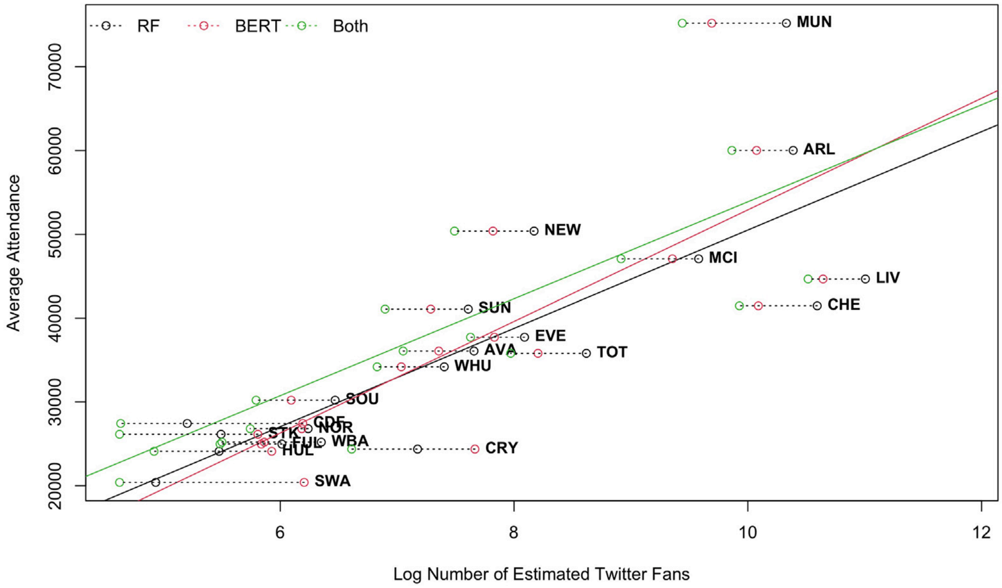

Fig. 2. Fans identified from sentiment and average attendance.
Notes: This Figure plots the logarithmized number of *fans* – identified by the Random Forest model (RF), the transformer model (BERT), or the overlap sample according to both models – and the average attendance at home games in season 2013/14. ARL: Arsenal, AVA: Aston Villa, CDF: Cardiff City, CHE: Chelsea, CRY: Crystal Palace, EVE: Everton, FUL: Fulham, HUL: Hull City, LIV: Liverpool, MCI: Manchester City, MUN: Manchester United, NEW: Newcastle United, NOR: Norwich City, SOU: Southampton, STK: Stoke City, SUN: Sunderland, SWA: Swansea City, TOT: Tottenham Hotspur, WBA: West Bromwich Albion, WHU: West Ham United.

2.2. Measuring emotional cues

Following Buraimo et al. (2020) we rely on the probability of each of the three outcomes in a football match – i.e., home win (H), draw (D), or away win (A) – at time t, denoted as p_t^H, p_t^D, and p_t^A respectively, for measuring emotional cues.

At first glance, it seems promising to take in-play betting data for deriving these probabilities on a minute-by-minute basis. In this regard, the most comprehensive data come from the *Betfair* betting exchange where offered prices evolve by betting market participants prepared to both buy and sell betting contracts. However, while some studies have shown that *Betfair*, or betting

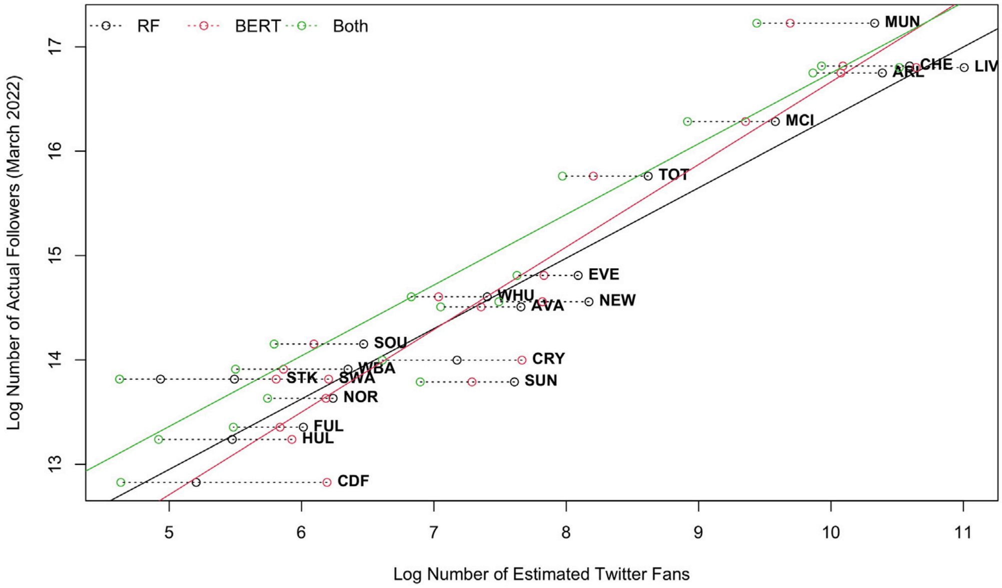

Fig. 3. Fans identified from sentiment and followers of team accounts.
Notes: This Figure plots the logarithmized number of *fans* — identified by the Random Forest model (RF), the transformer model (BERT), or the overlap sample according to both models — and the logarithmized number of followers of team accounts as of March 2022. ARL: Arsenal, AVA: Aston Villa, CDF: Cardiff City, CHE: Chelsea, CRY: Crystal Palace, EVE: Everton, FUL: Fulham, HUL: Hull City, LIV: Liverpool, MCI: Manchester City, MUN: Manchester United, NEW: Newcastle United, NOR: Norwich City, SOU: Southampton, STK: Stoke City, SUN: Sunderland, SWA: Swansea City, TOT: Tottenham Hotspur, WBA: West Bromwich Albion, WHU: West Ham United.

exchange, prices, accurately predict outcomes (Croxson and Reade, 2014), others have rejected the hypothesis of semi-strong market efficiency. For instance, Choi and Hui (2014) found that prices generally underreact to normal news and overreact to surprising news. Such market inefficiencies are also detected by Angelini et al. (2022). In summary, these findings question the overall suitability of using observable (*Betfair*) prices for predicting outcomes in our study.

In this study, we use in-play odds derived from an in-play model as proposed by Buraimo et al. (2020). The in-play model is built on pre-match closing odds in combination with over-under totals which reflect the strengths of teams in contention as well as other relevant factors such as current form of the teams and their most recent match results. By assuming an independent Poisson distribution for goals scored by both home and away teams and using the empirical goal distribution during EPL games it is possible to generate the probabilities for every scoreline for a given match and calculate the required outcome probabilities p_t^H, p_t^D, and p_t^A (for further details, see Appendix A).

For illustrating how both actual and simulated outcome probabilities develop during the course of a match, we take an example from the match between Crystal Palace and Liverpool on May 5 2014 (matchday 37 of 38). This was the first match after the home loss against Chelsea (mentioned in Section 2.1) and was as such also critical for the championship race that season. Liverpool was winning the match 3–0 until the 79th minute when goals by Delaney and Gayle (2) helped Crystal Palace to (unexpectedly) draw. The match ended 3–3 leaving Liverpool with hardly any chance of winning the title. Fig. 4 shows how actual and simulated probabilities developed during the course of this match. As could be expected, each goal by Liverpool is decreasing home win and draw probabilities while increasing away win probabilities (at *decreasing* margins). The opposite pattern can be observed for each goal scored by Crystal Palace, i.e., an increase in home win and draw probabilities as well as a fall in away win probabilities (at *increasing* margins). Note that changes in outcome probabilities are not only caused by goals scored (otherwise we would observe just flat lines between any goals scored). Overall, we only observe some minor differences between actual and simulated probabilities by visual inspection. In our analysis we use simulated instead of the actual probabilities for calculating our emotional cue measures for the reasons mentioned earlier.

Recall that *surprise* is a backward-looking measure which results from an outcome that contradicts anterior beliefs. Considering outcome probabilities as defined before and in line with Buraimo et al. (2020) we define surprise as:

$$Surprise_t = \sqrt{(p_t^H - p_{t-1}^H)^2 + (p_t^D - p_{t-1}^D)^2 + (p_t^A - p_{t-1}^A)^2}. \tag{1}$$

Shock is defined similarly, but with respect to the probabilities at the start of the match:

$$Shock_t = \sqrt{(p_t^H - p_0^H)^2 + (p_t^D - p_0^D)^2 + (p_t^A - p_0^A)^2}. \tag{2}$$

In contrast, however, *suspense* is a forward-looking measure which attempts to capture the impact of a goal scored in the next minute on either of the three outcome probabilities. We thus introduce p_{t+1}^{HS} and p_{t+1}^{AS} to denote the probability of the home and away teams

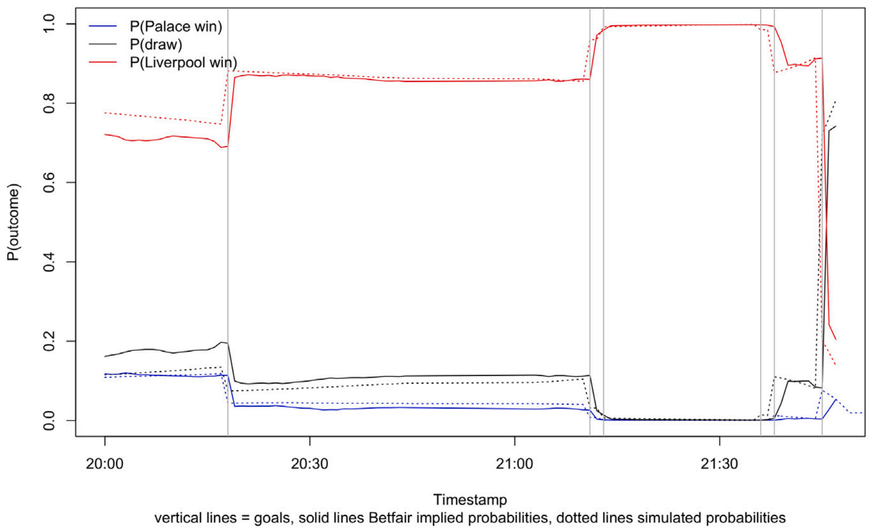

Fig. 4. Development of outcome probabilities during the course of a match.
Notes: This Figure plots the development of outcome probabilities during the course of the match between Crystal Palace and Liverpool on May 5 2014 (matchday 37 of 38). The outcome probabilities were either derived from *Betfair* exchange data sourced via *Fracsoft* (solid lines) or simulated with our in-play model (dotted lines) as described in Appendix A. Vertical lines indicate goals scored, i.e., 0–1 (Allen, 18'), 0–2 (Delaney own goal, 53'), 0–3 (Suarez, 55'), 1–3 (Delaney, 79'), 2–3 (Gayle, 81'), 3–3 (Gayle, 88').

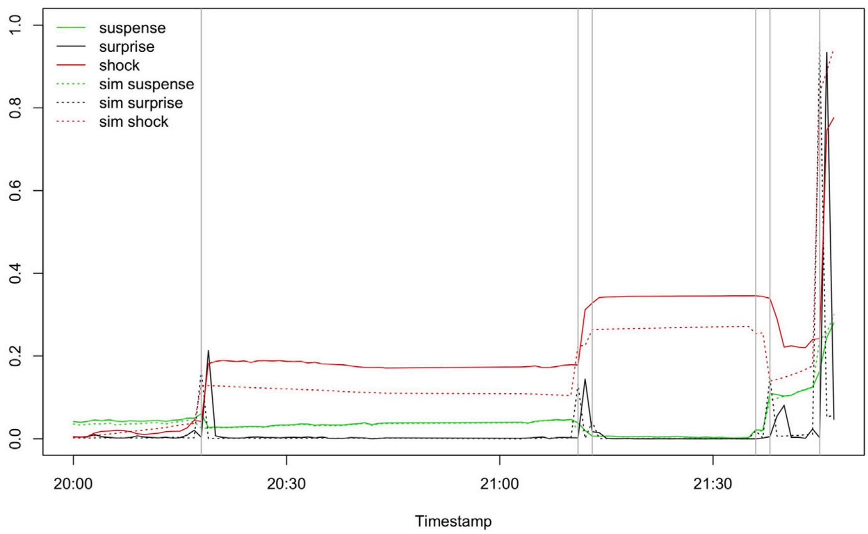

Fig. 5. Development of *surprise*, *shock*, and *suspense* during the course of a match.
Notes: This Figure plots the development of *surprise* (black), *shock* (red), and *suspense* (green) during the course of the match between Crystal Palace and Liverpool on May 5 2014 (matchday 37 of 38). *Surprise*, *shock*, and *suspense* were calculated from either *Betfair* exchange data sourced via *Fracsoft* (solid lines) or simulated odds (dotted lines) as described in Section 2.2 and Appendix A. Vertical lines indicate goals scored, i.e., 0–1 (Allen, 18'), 0–2 (Delaney own goal, 53'), 0–3 (Suarez, 55'), 1–3 (Delaney, 79'), 2–3 (Gayle, 81'), 3–3 (Gayle, 88').

scoring in the next minute. Then *suspense* is defined as:

$$Suspense_t = \left(\sum_{i \in H,D,A} p_{t+1}^{HS} \left[(p_{t+1}^{i} \left| p_{t+1}^{HS} \right.) - p_t^i \right]^2 + \sum_{i \in H,D,A} p_{t+1}^{AS} \left[(p_{t+1}^{i} \left| p_{t+1}^{AS} \right.) - p_t^i \right]^2 \right)^{1/2} \tag{3}$$

Taking the same example as before, Fig. 5 indicates how *shock*, *surprise*, and *suspense* develop during the course of the match. Overall, the observed patterns seem reasonable. While *suspense* gradually *decreases* up to the 79th minute when Crystal Palace scored to make the scoreline 1–3, it substantially *increases* particularly after the third goal scored by Crystal Palace. Likewise, *shock* and *surprise* are mainly driven by the goals scored. More broadly speaking, *suspense* commonly reflects an upward trend over time up to the point when a match is (most likely) decided. In contrast, the pattern of *surprise* is spiky and mainly depends on goals scored.

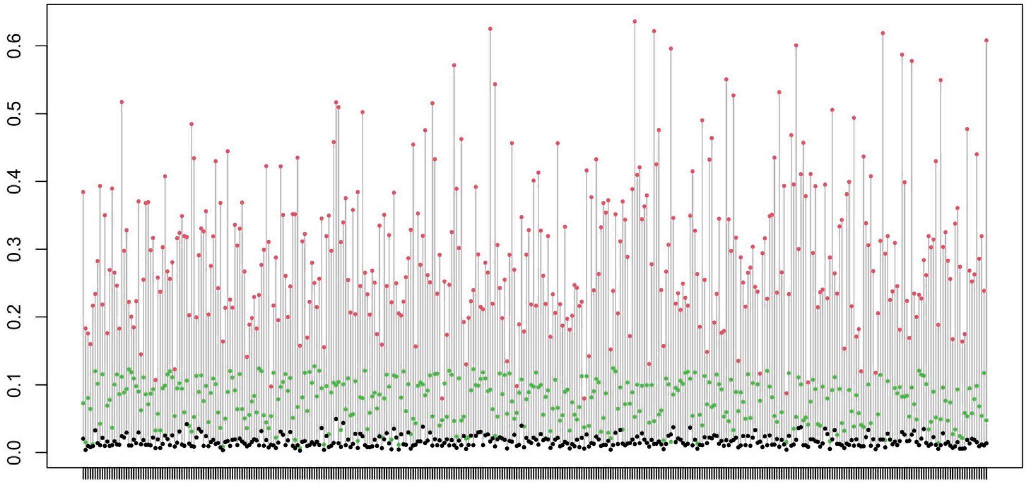

Fig. 6. Mean shock, surprise, and suspense per match.
Notes: This Figure plots the mean *surprise* (black), *shock* (red), and *suspense* (green) per match for all 380 matches played in season 2013/2014 calculated from simulated odds as described in Section 3 and Appendix A.

Finally, it is worth noting that we not only observe variation in *surprise*, *shock*, and *suspense within* a match but also *between* matches (see Fig. 6). This must be considered in our empirical model.

2.3. Empirical model

In this study, we intend to model the extent to which emotional cues from football experience, i.e., *surprise*, *shock*, and *suspense*, provoke measurable behavioral responses. As such, the number of Tweets in a given minute t of match i serves as the dependent variable y_{it} in our empirical model:

$$y_{it} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 surprise_{it} + \beta_3 shock_{it} + \beta_4 suspense_{it} + \beta_5 X_{it} + \beta_6 home_i + \delta_t + \gamma_i + \nu_c + u_{it}. \quad (4)$$

In our main specification, we control for lagged number of Tweets y_{it-1} and whether the Tweet is home team-related. Moreover, in order to pick up any differences between minutes played as well as across matches and teams, we control for minute fixed effects δ_t, match fixed effects γ_i, and team fixed effects ν_c. As a robustness check, we further control for a set of in-play events X_{it} like goals scored, shots, corners, cards, substitutions, as well as total goals scored. Since we expect certain match-specific dynamics and as such within-match correlation, standard errors are clustered by match.

Our regressions are run separately for Twitter users that we have identified as *fans*, *haters*, and *neutrals*. Since a match involves two teams, we have to consider both home team- and away team-related Tweets leaving us with two observations/counts per match-minute in all regression models. For example, in the *fan* regressions the first count covers any Tweet released by an identified *fan* of the home team irrespective of whether the Tweet uses a hashtag of the home team, the away team, or both types of hashtags. The second count covers any Tweet released by an identified *fan* of the away team irrespective of whether the Tweet uses a hashtag of the away team, the home team, or both. This works accordingly in the *hater* regressions. It is worth noting that about one out of ten Tweets by both *fans* and *haters* was posted with a hashtag of the opposing team, reinforcing the relevance of our procedure. Finally, the counts of Tweets in the *neutral* regressions are made up of both *neutral* users plus *fans/haters* of teams other than the two that are participating in the match, who tweet using any hashtag associated with one of the teams playing. Overall, we observe a remarkable variation in the number of Tweets by *fans*, *haters*, and *neutrals* across the matches in our sample (see in Fig. 7).[14]

3. Results

Table 1 provides an overview of our regression results separated for *fans*, *haters*, and *neutrals*. For each of these groups we present results from three different specifications, i.e., based on the identification of *fans*, *haters*, and *neutrals* according to the Random Forest model (RF), the transformer model (BERT), and the overlap sample from both models (overlap). Given the comparably small number of identified *haters* in the 'overlap sample', findings of the corresponding regressions have to be treated with caution. All regressions control for the logarithmized lagged number of Tweets, home team-related Tweets, as well as minute, match, and team fixed effects.

[14] The corresponding Figures according to the transformer model (BERT) and the overlap sample identified by both models can be found in Appendix B Figs. B3 and B4.

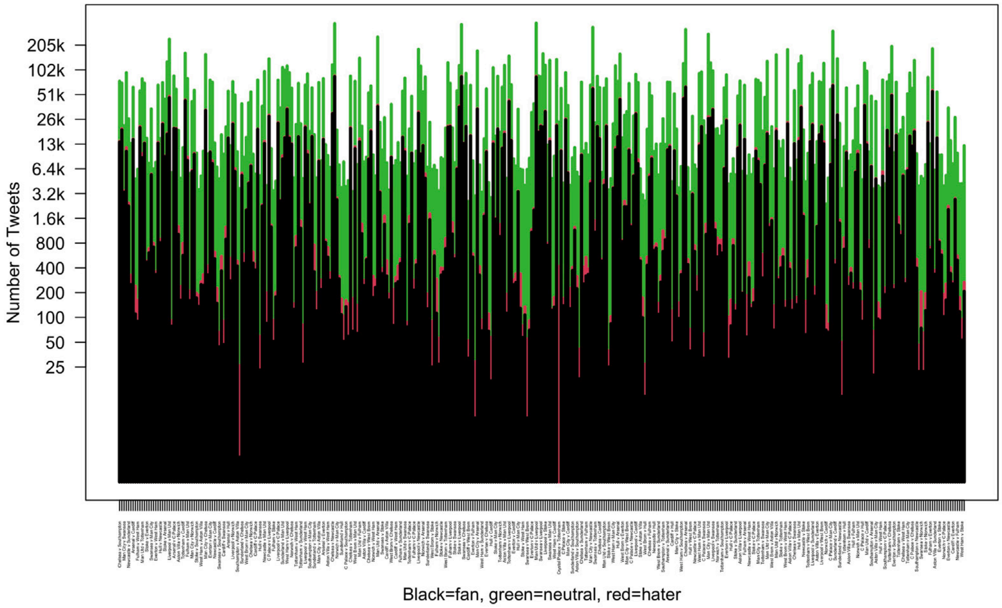

Fig. 7. Tweets per match and type of user according to the RF model.
Notes: This Figure plots the number of Tweets per match by *neutrals* (green), *haters* (red), and *fans* (black) identified by the Random Forest model (RF) for all 380 matches played in season 2013/2014.

Table 1
Results for Tweets by *fans*, *haters*, and *neutrals*.

	Dependent variable: log number of Tweets by...								
	Fans			*Haters*			*Neutrals*		
	(1) RF	(2) BERT	(3) Overlap	(4) RF	(5) BERT	(6) Overlap	(7) RF	(8) BERT	(9) Overlap
Surprise	2.040***	2.145***	1.700***	1.863***	1.712***	0.597	0.508***	0.610***	0.655***
	(0.520)	(0.582)	(0.490)	(0.486)	(0.574)	(0.447)	(0.114)	(0.099)	(0.093)
Shock	2.386***	2.809***	2.105***	4.433***	3.637***	2.806***	0.885***	0.848***	0.750***
	(0.313)	(0.342)	(0.298)	(0.366)	(0.344)	(0.308)	(0.124)	(0.130)	(0.131)
Suspense	−7.266***	−7.910***	−6.837***	−9.508***	−7.571***	−7.468***	−1.975***	−2.052***	−1.856***
	(1.220)	(1.301)	(1.133)	(1.507)	(1.320)	(1.248)	(0.506)	(0.492)	(0.448)
Lag Tweets	✓	✓	✓	✓	✓	✓	✓	✓	✓
Home team	✓	✓	✓	✓	✓	✓	✓	✓	✓
Minute FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Match FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Team FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	67,233	67,233	67,233	67,233	67,322	67,322	67,322	67,233	67,233
R^2	0.569	0.539	0.559	0.431	0.457	0.399	0.803	0.834	0.866
Adjusted R^2	0.566	0.536	0.556	0.427	0.454	0.394	0.802	0.833	0.865
Residual s.e.	6.228	6.430	6.951	6.381	6.355	6.033	1.450	1.316	1.167
Degrees of free.	66,742	66,742	66,742	66,742	66,742	66,742	66,742	66,742	66,742

Notes: This Table provides an overview of the effects of emotional cues on the number of Tweets across Twitter users. The logarithmized number of Tweets associated either with the corresponding home team or away team by *fans* (Columns 1–3), *haters* (Columns 4–6), and *neutrals* (Columns 7–9) serves as dependent variable in the models. All models include the logarithmized lagged number of Tweets, home team as well as minute, match, and team fixed effects. Specifications in Columns (1), (4), and (7) are based on the identification of *fans*, *haters*, and *neutrals* according to the Random Forest model (RF). Specifications in Columns (2), (5), and (8) are based on the identification of *fans*, *haters*, and *neutrals* according to the transformer model (BERT). Specifications in Columns (3), (6), and (9) are based on the identification of *fans*, *haters*, and *neutrals* in the overlap sample according to both models (overlap). Standard errors (in brackets) are clustered by match. Significance levels are $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Since we control for lagged number of Tweets and excluded extra time, these regressions are based on 89 min for 380 games. As we make use of both home team- and away team-related Tweets, we end up with 67,233 minute-game observations.[15]

[15] Note, we miss a small number of minute-game-observations. As such, we end up with 67,233 instead of 67,640 minute-game observations (i.e., 89 min × 380 games × 2 teams).

Overall, we find that emotional cues significantly influence the number of Tweets in a given minute. While *surprise* and *shock increase* the number of Tweets, *suspense reduces* the number of Tweets. As intuitively expected, effect sizes for all emotional cues are the smallest for *neutrals*. Moreover, visual inspection of all coefficients suggests a tendency that effect sizes of *shock* and *suspense* are larger for *haters* than for *fans*. However, given the structure of our complex data we are not able to directly support this observation with a formal test. In principle, our findings are very similar across the different types of models for identifying *fans*, *haters*, and *neutrals*. However, the estimate for *surprise* in the *haters* 'overlap sample' is not precise enough to be significant at conventional levels.

Controlling for in-play events – i.e., dummy variables indicating goal, lagged goal (in the minute before), shot, shot hit goalframe, corner, yellow card, red card, or substitution, as well as total goals scored – reduces effect sizes for all three cue measures (see Appendix B, Table B1). This could be expected by construction, because simulating in-play odds (by us and the bookmakers) involves exploiting partly the same information during the course of the game about, for instance, goals scored or red cards received (see Section 2.2 and Appendix A.2). Nevertheless, even though reduced in size, the effects of *surprise* (though only for the *neutrals* sample), as well as *shock*, and *suspense* remain significant at the conventional level suggesting that our cues pick up some elements of emotions beyond just scoring a goal, for instance.

Despite the concerns when using bookmaker odds (see Section 2.2) we are able to test whether our findings are driven by using simulated odds. As shown in Table B2 in Appendix B, the results look similar when cues are based on outcome probabilities derived from *Betfair* exchange data sourced via *Fracsoft*. The only main difference is the size of the coefficient for *surprise* which is about four times as large compared to our main specification in Table 1. A possible reason could be that the effect of *surprise* takes some time to unfold. As such, it would be better picked up using real odds which commonly reflect a short delay for updating (see Figs. 4 and 5).[16]

Because the dependent variable in all models is a count variable and we partly observe large numbers of zeros in our data (i.e., minutes where no Tweets are posted) – particularly for the *haters* samples – we also explore the sensitivity of our findings regarding the type of estimator used. Table B3 in Appendix B displays the results from our main specifications using Poisson regressions instead of OLS. It is important to note that these results are not directly comparable to our main specifications because the *pgml*-command in *R* does not allow for clustering standard errors, for instance. Moreover, we use lagged number of Tweets instead of the logarithmized lagged number of Tweets in order to better consider the nature of the dependent variable in these count data models. Overall, and despite these differences between specifications, our main findings hold.

In a further robustness check, we introduce the element of 'Tweet relevance' by considering a joint measure of the number of Tweets by *fans*, *haters*, and *neutrals* plus the corresponding Retweets. As shown in Table B4 in Appendix B, the estimates are very similar compared with our main specifications in Table 1. However, the effects of *surprise* for *neutrals* (in addition to the *haters* 'overlap sample') are not precise enough to be significant in this specification.

Finally, in order to further explore the relevance of a particular course of the match, we add variables measuring whether the favorite (or hated) team is currently winning or losing along with the corresponding interactions between winning/losing and our emotional cue measures. Following this approach and given the temporal structure of all measures, the interpretation is as follows: if, for instance, a team scores and goes ahead, that effect on *surprise* is part of the *surprise*-winning interaction. If a team concedes and goes behind, that effect on *surprise* is part of the *surprise*-losing interaction. If a team scores (or concedes) an equalizer, that effect is covered in the normal *surprise* coefficient.

From our results in Table 2 (for *fans*) and Table 3 (for *haters*), the main findings remain for *surprise* and *shock*. Interestingly, however, while both winning and losing increase the number of Tweets, all interactions between winning/losing and our cue variables are negative and most often strongly significant. This suggests that all cues unfold their largest *immediate* effects when the match is currently a tie or has just become a tie. In other words, our findings suggest that goal-induced *immediate* effects from *surprise* and *shock* on Twitter activity are the largest, when the favorite (or hated) team either scores or concedes an equalizer.

A closer look into the findings for *fans* reveals, that the effect of *surprise* when winning is much smaller compared to the effect of *surprise* when losing. The (negative) estimate of the interaction when winning and the (positive) estimate for *surprise* are of similar size or even cancel out. In contrast, the (negative) estimate of the interaction when losing is much smaller and hardly significant at conventional levels (see Columns (1)-(3) in Table 2). For *haters*, we observe the opposite pattern (see Columns (1)-(3) in Table 3). Taking these observations together we find that goal-induced *immediate* effects from *surprise* on Twitter activity are particularly small (though existent) when the probability of the 'desired' result – i.e., winning of the favorite team (for *fans*) and losing of the hated team (for *haters*) – increases.

Regarding *suspense*, we observe that the negative effect only remains when losing or winning. This suggests that *suspense* does not unfold any negative effect when the match is currently a tie. Moreover, for *haters*, the pattern of the effects of *suspense* is similar to the pattern of the effects of *surprise*, i.e., Twitter activity is most reduced by *suspense* when the hated team is losing. For *fans*, however, we do not observe such asymmetries.

[16] As mentioned earlier, we refrain from further exploring any lagged effects in our setting given econometric concerns caused by the temporal structure of the data with many measurement points.

Table 2
Results for Tweets by *fans* considering winning and losing.

	Dependent variable: log number of Tweets by *fans*								
	Surprise interaction			Shock interaction			Suspense interaction		
	(1) RF	(2) BERT	(3) Overlap	(4) RF	(5) BERT	(6) Overlap	(7) RF	(8) BERT	(9) Overlap
Surprise	3.961***	4.952***	3.893***	1.788***	1.860***	1.473***	1.791***	1.870***	1.475***
	(0.934)	(1.039)	(0.915)	(0.524)	(0.581)	(0.487)	(0.522)	(0.583)	(0.485)
Shock	1.464***	1.789***	1.303***	3.344***	3.980***	3.056***	1.129***	1.437***	0.998***
	(0.376)	(0.414)	(0.350)	(0.530)	(0.578)	(0.502)	(0.392)	(0.435)	(0.363)
Suspense	−3.974***	−4.343***	−4.025***	−7.447***	−8.339***	−7.231***	2.120	2.274	1.634
	(1.335)	(1.414)	(1.202)	(1.542)	(1.680)	(1.373)	(1.807)	(1.878)	(1.686)
Winning	1.168***	1.203***	1.081***	1.288***	1.300***	1.318***	1.900***	1.841***	1.709***
	(0.144)	(0.144)	(0.135)	(0.215)	(0.215)	(0.201)	(0.259)	(0.261)	(0.241)
Losing	0.249*	0.388***	0.172	0.837***	1.082***	0.571***	1.025***	1.336***	0.915***
	(0.133)	(0.141)	(0.124)	(0.212)	(0.227)	(0.198)	(0.229)	(0.254)	(0.220)
Cue × winning	−4.325***	−5.865***	−4.661***	−1.810***	−2.036***	−2.035***	−10.010***	−8.724***	−8.670***
	(1.198)	(1.245)	(1.082)	(0.630)	(0.665)	(0.561)	(2.682)	(2.690)	(2.503)
Cue × losing	−1.482	−2.469*	−1.857	−2.887***	−3.420***	−2.332***	−9.966***	−12.838***	−9.795***
	(1.262)	(1.338)	(1.224)	(0.664)	(0.711)	(0.627)	(2.386)	(2.631)	(2.375)
Lag Tweets	✓	✓	✓	✓	✓	✓	✓	✓	✓
Home team	✓	✓	✓	✓	✓	✓	✓	✓	✓
Minute FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Match FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Team FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	67,233	67,233	67,233	67,233	67,322	67,322	67,322	67,233	67,233
R^2	0.571	0.540	0.561	0.571	0.541	0.561	0.571	0.540	0.561
Adjusted R^2	0.568	0.537	0.558	0.568	0.537	0.558	0.568	0.537	0.558
Residual s.e.	6.217	6.421	5.941	6.216	6.420	5.941	6.210	6.420	5.940
Degrees of free.	66,738	66,738	66,738	66,738	66,738	66,738	66,738	66,738	66,738

Notes: This Table provides an overview of the effects of emotional cues on the number of Tweets for *fans*. The logarithmized number of Tweets associated either with the corresponding home team or away team serves as dependent variable in the models. All models include the logarithmized lagged number of Tweets, home team as well as minute, match, and team fixed effects. In contrast to results presented in Table 1, all specification also include variables which measure whether the favorite team is currently winning or losing as well as the corresponding interactions with *surprise* (Columns 1–3), *shock* (Columns 4–6), and *suspense* (Columns 7–9). Specifications in Columns (1), (4), and (7) are based on the identification of *fans* according to the Random Forest model (RF). Specifications in Columns (2), (5), and (8) are based on the identification of *fans* according to the transformer model (BERT). Specifications in Columns (3), (6), and (9) are based on the identification of *fans* in the overlap sample according to both models (overlap). Standard errors (in brackets) are clustered by match. Significance levels are $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

4. Discussion and conclusions

By analyzing 19 m Tweets in combination with in-play information for overall 380 games played in the English Premier League we analyze whether and how emotional cues influence activity on Twitter. In contrast to other popular social media platforms like Facebook or Instagram where algorithms mainly determine the news feeds, news on Twitter are chronological.[17] Moreover, as a microblogging platform, Twitter forces concise communication by only allowing for a maximum of 140 characters per Tweet (during our observation window in 2013/14). This enables users to quickly react to news/events and express opinions and sentiments (Stiefel and Vivés, 2022). Taking all this together, Twitter became a popular medium for empirical analysis on sentiment and human behavior in other disciplines like finance (see, for instance, Bartov et al. (2018), Grebe et al. (2024), Gu and Kurov (2020), or Puklavec et al. (2023)). Given the popularity of Twitter in sports already at the time when our data was gathered,[18] we argue that Twitter is the first best medium to use for our research purpose. In particular, this setting in combination with our proposed approach of exploiting sentiment expressed in Tweets provides the unique opportunity to explore the behavioral response to emotional cues separately for *fans*, *haters*, and *neutrals* in a comprehensive manner.

A potential criticism of our data is that it is a number of years old, hailing from the 2013/2014 season. We would stress that taking data from such a period allows us to address our research question at a time when chatbots and other potentially manipulating techniques or institutions did not play a major role. Moreover and importantly, even though the way Twitter is used in society has changed over time, we do not see any reason to believe that Twitter activity as a response to emotional cues from sports should have changed systematically. As such, we argue that the data at hand allow for a valid and timely empirical test of the effects of interest. At the same time, however, Twitter data (in the past and in the present) suffer from a lack of reliable sociodemographic information and location data of users. People regularly use fake information or just do not reveal this information. As such, we can

[17] In the time since our observation window of 2013/14 Twitter, now known as X, has developed a more algorithmic method to determine news feeds dependent on user preferences, if the user so desires. At the time of our sample, however, tweets would appear in chronological order.

[18] E.g., 50 percent of all Tweets in 2013 about TV in the US were sports event-related (Nielsen, 2014).

Table 3
Results for Tweets by *haters* considering winning and losing.

	Dependent variable: log number of Tweets by *haters*								
	Surprise interaction			Shock interaction			Suspense interaction		
	(1) RF	(2) BERT	(3) Overlap	(4) RF	(5) BERT	(6) Overlap	(7) RF	(8) BERT	(9) Overlap
Surprise	4.819***	4.074***	3.771***	1.528***	1.412**	0.317	1.545***	1.419**	0.316
	(0.940)	(0.967)	(0.810)	(0.492)	(0.566)	(0.441)	(0.492)	(0.564)	(0.439)
Shock	3.095***	2.374***	1.602***	5.284***	4.131***	3.224***	2.772***	2.090***	1.326***
	(0.409)	(0.400)	(0.335)	(0.592)	(0.570)	(0.519)	(0.412)	(0.412)	(0.345)
Suspense	−4.793***	−3.075***	−3.288***	−8.775***	−6.267***	−6.161***	1.387	2.309	2.159
	(1.649)	(1.450)	(1.279)	(1.889)	(1.704)	(1.512)	(1.989)	(1.859)	(1.733)
Winning	0.868***	0.958***	0.919***	1.357***	1.298***	1.583***	1.452***	1.500***	1.447***
	(0.134)	(0.143)	(0.134)	(0.195)	(0.216)	(0.198)	(0.225)	(0.245)	(0.230)
Losing	1.184***	0.956***	0.939***	1.471***	1.239***	0.788***	2.054***	1.693***	1.650***
	(0.140)	(0.147)	(0.135)	(0.187)	(0.207)	(0.195)	(0.217)	(0.245)	(0.228)
Cue × winning	−2.828***	−2.850**	−4.060***	−2.911***	−2.228***	−3.014***	−7.278***	−6.973***	−7.129***
	(1.242)	(1.309)	(1.200)	(0.660)	(0.663)	(0.626)	(2.632)	(2.657)	(2.504)
Cue × losing	−6.067***	−4.346***	−5.371***	−2.531***	−2.143***	−0.984	−12.785***	−10.528***	−10.422***
	(1.253)	(1.195)	(1.045)	(0.597)	(0.635)	(0.626)	(2.425)	(2.543)	(2.480)
Lag Tweets	✓	✓	✓	✓	✓	✓	✓	✓	✓
Home team	✓	✓	✓	✓	✓	✓	✓	✓	✓
Minute FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Match FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Team FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	67,233	67,233	67,233	67,233	67,322	67,322	67,322	67,233	67,233
R^2	0.432	0.458	0.400	0.433	0.459	0.400	0.433	0.459	0.400
Adjusted R^2	0.428	0.454	0.395	0.428	0.455	0.396	0.428	0.455	0.395
Residual s.e.	6.374	6.349	6.027	6.373	6.349	6.026	6.373	6.348	6.027
Degrees of free.	66,738	66,738	66,738	66,738	66,738	66,738	66,738	66,738	66,738

Notes: This Table provides an overview of the effects of emotional cues on the number of Tweets for *haters*. The logarithmized number of Tweets associated either with the corresponding home team or away team serves as dependent variable in the models. All models include the logarithmized lagged number of Tweets, home team as well as minute, match, and team fixed effects. In contrast to results presented in Table 1, all specification also include variables which measure whether the disliked team is currently winning or losing as well as the corresponding interactions with *surprise* (Columns 1–3), *shock* (Columns 4–6), and *suspense* (Columns 7–9). Specifications in Columns (1), (4), and (7) are based on the identification of *haters* according to the Random Forest model (RF). Specifications in Columns (2), (5), and (8) are based on the identification of *haters* according to the transformer model (BERT). Specifications in Columns (3), (6), and (9) are based on the identification of *fans* in the overlap sample according to both models (overlap). Standard errors (in brackets) are clustered by match. Significance levels are $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

neither explore effect heterogeneity with regard to, for instance, gender and nationality, nor can we explore the role of contextual factors like weather conditions at the location where a Tweet was released.

Overall, we find that both *surprise* and *shock* immediately increase the number of Tweets, while *suspense* (on average) decreases the number of Tweets. In this regard, the negative effect of *suspense* on Twitter activity is suggestive of individuals being ‘caught in the moment’ and as such paying more attention to the match itself. This proposition is fully backed up by studies exploring the demand for sports telecasts which unambiguously reveal a positive effect of *suspense* on viewing figures (see, for instance, Buraimo et al., 2020 or Richardson et al., 2023). Moreover, it is in line with a recent study which finds *suspense* to reduce alcohol consumption during a match (Fischer et al., 2023).

Moreover, as could be expected, any response to emotional cues is smallest for *neutrals* compared to *fans* and *haters*. Further analysis reveals that goal-induced *immediate* effects from *surprise* and *shock* on Twitter activity are the largest, when the favorite (or hated) team either scores or concedes an equalizer. A closer look into these findings reveals some interesting similarities in behavior between *fans* and *haters*. Goal-induced *immediate* effects from *surprise* are particularly small when the probability of the ‘desired’ result — i.e., winning of the favorite team (for *fans*) and losing of the hated team (for *haters*) — increases. On the one hand, this observation could be suggestive of some sort of extraordinary celebration taking place in the minute of the ‘desired’ goal, thus eventually postponing Twitter activity. More importantly, however, this observation adds further credibility to our identification exercise because one would expect that both *fans* and *haters* respond in a similar way when the probability of winning of the favorite team (for *fans*) and losing of the hated team (for *haters*) increases.

These findings could inform the literature in three ways. First, we follow the call by Loewenstein (2000) and provide new evidence of how *immediate* emotions influence *immediate* human behavior. Our setting is highly relevant given that the global sports market is nowadays a multi-billion dollar business with a value of more than $388 billion in 2020 (TBRC, 2023). Moreover, it seems particularly promising since professional sports is frequently regarded as *the* emotions lab and Tweeting is an easy-to-measure and straight forward activity for millions of people around the world. Second, as could be seen in our analysis, *fans*, *haters*, and *neutrals* respond to emotional cues partly differently. From a managerial perspective this might be relevant to consider when implementing personalized forms of communication by clubs and sponsors during the course of a match. Third, by looking at behavioral responses to the temporal resolution of uncertainty during the course of a game, we contribute to the literature testing the uncertainty-of-outcome hypothesis in sports by using more fine-grained data. In fact, our findings suggest that entertainment utility is driven by both elements which gain in (lagged) *certainty* (such as surprise and shock) as well as elements which gain in *uncertainty* (such as suspense).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: James Reade, Dooruj Rambacussing and Giambattista Rossi acknowledge that part of the data used in this paper was provided by Twitter through their #DataGrants scheme.

Data availability

We are sharing the data on a Github site: https://github.com/philiprami/surprise_suspense_sentiment_shock.

Acknowledgments

We thank Lukas E. G. Fischer and Michael Nagel for their assistance in the development of the emotional cue measures. A previous version of this paper was presented in the 'Reading Online Sport Economics' Seminar (2022), in the workshop 'Machine Learning meets Quantitative Social Sciences' in Tübingen (2022), in the 'Gambling Symposium' in Hohenheim (2023), at the annual meetings of the European Association for Sport Management in Innsbruck (2022) and the German Association of Sport Economics and Sport Management in Innsbruck (2023), at the Football & Finance Conference in Paderborn (2023), as well as in seminars at the Universities of Aveiro (2022), St. Gallen (2022), and Zurich (2022). We thank participants as well as Alex Krumer, Detmar Meurers, Ignacio Palacios-Huerta, and Thomas Peeters for their helpful comments and suggestions. Moreover, we thank the Co-Editor, Scott Adams, as well as two anonymous referees for their guidance and help in improving the manuscript. James Reade, Dooruj Rambacussing and Giambattista Rossi acknowledge that part of the data used in this paper was provided by Twitter through their #DataGrants scheme. Moreover, this work was supported with a Research Grant by the School of Business, Economics and Informatics (Birbeck University of London). The funders had no role in the study design, the data collection and analysis, or the preparation of the manuscript. The authors do not have a financial interest in the topic of this paper. There are no conflicts of interest. All errors are our own.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jebo.2024.04.018.

References

Angelini, G., De Angelis, L., Singleton, C., 2022. Informational efficiency and behaviour within in-play prediction markets. Int. J. Forecast. 38 (1), 282–299.
Babad, E., Katz, Y., 1991. Wishful thinking—against all odds. J. Appl. Soc. Psychol. 21 (23), 1921–1938.
Bartov, E., Faurel, L., Mohanram, P., 2018. Can Twitter help predict firm-level earnings and stock returns? Account. Rev. 93, 25–57.
Bizzozero, P., Flepp, R., Franck, E., 2016. The importance of suspense and surprise in entertainment demand: Evidence from Wimbledon. J. Econ. Behav. Organ. 130, 47–63.
Buraimo, B., Forrest, D., McHale, I., Tena, J., 2020. Unscripted drama: Soccer audience response to suspense, surprise, and shock. Econ. Inq. 58 (2), 881–896.
Caplin, A., Leahy, J., 2001. Psychological expected utility theory and anticipatory feelings. Q. J. Econ. 116 (1), 55–79.
Choi, D., Hui, S., 2014. The role of surprise: Understanding overreaction and underreaction to unanticipated events using in-play Soccer betting market. J. Econ. Behav. Organ. 107, 614–629.
Croxson, K., Reade, J., 2014. Information and efficiency: Goal arrival in Soccer betting. Econom. J. 124 (575), 62–91.
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
Dillenberger, D., 2010. Preferences for one-shot resolution of uncertainty and Allais-type behavior. Econometrica 78 (6), 1973–2004.
Ely, J., Frankel, A., Kamenica, E., 2015. Suspense and surprise. J. Polit. Econ. 123 (1), 215–260.
Fischer, L.E., Nagel, M., Kelava, A., Pawlowski, T., 2023. Celebration beats frustration – emotional cues and alcohol use during Soccer matches. http://dx.doi.org/10.2139/ssrn.4569227, SSRN Working Paper.
Grebe, M., Kandemir, S., Tillmann, P., 2024. Uncertainty about the war in Ukraine: Measurement and effects on the German economy. J. Econ. Behav. Organ. 217, 493–506.
Gu, C., Kurov, A., 2020. Informational role of social media: Evidence from Twitter sentiment. J. Bank. Financ. 121 (105969).
Gul, F., 1991. A theory of disappointment aversion. Econometrica 59 (3), 667–686.
Jiwa, M., Cooper, P., Chong, T.T.-J., Bode, S., 2021. Choosing increases the value of non-instrumental information. Sci. Rep. 11 (8780).
Kaplan, S., 2021. Entertainment utility from skill and thrill. http://dx.doi.org/10.2139/ssrn.3888785, SSRN Working Paper.
Kőszegi, B., Rabin, M., 2009. Reference-dependent consumption plans. Amer. Econ. Rev. 99 (3), 909–936.
Kreps, D., Porteus, E., 1978. Temporal resolution of uncertainty and dynamic choice theory. Econometrica 46 (1), 185–200.
Liu, X., Shum, M., Uetake, K., 2021. Passive vs. active attention to baseball telecasts: Implications for content (re-)design. http://dx.doi.org/10.2139/ssrn.3717894, SSRN Working Paper.
Loewenstein, G., 2000. Emotions in economic theory and economic behavior. Amer. Econ. Rev. 90 (2), 426–432.
Lucas, G.M., Gratch, J., Malandrakis, N., Szablowski, E., Fessler, E., Nichols, J., 2017. GOAALLL: Using sentiment in the world cup to explore theories of emotion. Image Vis. Comput. 65, 58–65.
Neale, W.C., 1964. The peculiar economics of professional sports. Q. J. Econ. 78 (1), 1–14.

Nielsen, 2014. Sports Fans Amplify the Action Across Screens. Media Release, https://www.nielsen.com/insights/2014/sports-fans-amplify-the-action-across-screens/1.

Palacios-Huerta, I., 1999. The aversion to the sequential resolution of uncertainty. J. Risk Uncertain. 18 (3), 249–269.

Pawlowski, T., Nalbantis, G., Coates, D., 2018. Perceived game uncertainty, suspense and the demand for sport. Econ. Inq. 56 (1), 173–192.

Puklavec, Z., Kogler, C., Stavrova, O., Zeelenberg, M., 2023. What we tweet about when we tweet about taxes: A topic modelling approach. J. Econ. Behav. Organ. 212, 1242–1254.

Raney, A.A., 2018. Why we watch and enjoy mediated sports. In: Raney, A.A., Bryant, J. (Eds.), Handbook of Sports and Media. Lawrence Erlbaum Associates, New York, NY, pp. 313–329.

Richardson, T., Nalbantis, G., Pawlowski, T., 2023. Emotional cues and the demand for televised sports: Evidence from the UEFA champions league. J. Sports Econ. 24 (8), 993–1025.

Rottenberg, S., 1956. The baseball players' labor market. J. Polit. Econ. 64 (3), 242–258.

Simonov, A., Ursu, R.M., Zheng, C., 2022. Suspense and surprise in media product design: Evidence from twitch.tv. J. Mar. Res. 60 (1), 1–24.

Stiefel, M., Vivés, R., 2022. 'Whatever it takes' to change belief: Evidence from Twitter. Rev. World Econ. 158, 715–747.

TBRC, 2023. Sports Market 2023. Report by the Business Research Company, https://www.thebusinessresearchcompany.com/report/sports-market.

Zillmann, D., Cantor, J.R., 1972. Directionality of transitory dominance as a communication variable affecting humor appreciation. J. Personal. Soc. Psychol. 24 (2), 191–198.