Department of Economics

Group for Economic Analysis at Reading (GEAR)

Exploring Entertainment Utility from Football Games

by Tim Pawlowski, Dooruj Rambaccussing, Philip Ramirez, J. James Reade and Giambattista Rossi

Discussion Paper No. 2023-13

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

www.reading.ac.uk

Exploring entertainment utility from football games*

Tim Pawlowski † Dooruj Rambaccussing ‡ Philip Ramirez §

J. James Reade¶ Giambattista Rossi‖

July 7, 2023

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 participants involved in an event. By analyzing 19m 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 desireable 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.

JEL Classification: C10, D91, L83

Keywords: Suspense, surprise, entertainment utility, football, tweets

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

†University of Tübingen, Wilhelmstraße 124, DE 72074 Tübingen, tim.pawlowski@uni-tuebingen.de

‡University of Dundee, 1–3 Perth Road, d.rambaccussing@dundee.ac.uk

§University of Reading, Whiteknights, PO Box 217, p.ramirez@pgr.reading.ac.uk

¶University of Reading, Whiteknights, PO Box 217, j.j.reade@reading.ac.uk

‖Birkbeck University of London, Malet Street, Bloomsbury, g.rossi@bbk.ac.uk

1 Introduction

Modelling 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 modelling 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 modelled 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 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 modelling 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.

Overall, these studies find that suspense and — at least to some extent — also surprise and shock are important drivers of demand. Detailed findings, however, 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.

In contrast, skill primarily impacts viewership on the extensive margin, i.e., across games (Kaplan, 2021). (iv) Spectators have a higher probability to turn on games featuring less popular players/teams if they're nearing the end *and* exhibiting sufficiently high suspense (Kaplan, 2021). Finally, (v) postseason games amplify the effects of suspense and surprise while 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 behaviour, and in particular behaviour on Twitter, where individuals decide whether to send a Tweet.[1] 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

[1]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.

between different types of individuals. We start by generating sentiment scores for each Tweet using a random forest estimator trained on data from Stanford's Sentiment Tree Bank. 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 backward-looking measures *increase* the number of Tweets, *suspense* as a forward-looking measure *decreases* the number of Tweets. The latter could be explained by individuals being 'caught in the moment' probably leaving no time to tweet. As could be expected, any response to emotional cues is smallest for *neutrals*. Interestingly, however, *haters* respond more strongly than *fans* to such cues. Further analysis suggests that *goal-induced* effects from *surprise* and *shock* on Twitter activity are the largest, when the favorite (or hated) team either scores or concedes an equaliser. Moreover, we observe asymmetries particularly regarding the response to *suspense*, i.e., very suspenseful moments during a match when 'their' team is losing increase the number of Tweets by *fans* but not by *haters* while the corresponding effects from *suspense* remain negative when 'their' team is winning.

We contribute to 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 modelling 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.[2] Third, by looking at behavioral responses to the temporal resolution of uncertainty during the course of a game, we offer a different and fine-grained type of empirical test for the well-known uncertainty-of-outcome hypothesis in sports.[3] 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 labour market interventions in professional sport leagues around the globe (see, for instance, Pawlowski et al. (2018)).[4] 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 particular, we argue that 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 Fischer et al. (2023) who find that *suspense* reduces alcohol purchases in the stadium during a match.

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

[3] 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.

[4] 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 (out of 64) FIFA World Cup games in 2014, i.e., Vegas betting odds in order to measure differences between predicted and actual scores for the two teams in contention, 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 can hardly reveal any credible and robust evidence on the relation of interest. 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.

In Section 2 we introduce our data and methodology, in Section 3 we present our results, and Section 4 concludes.

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.[5] This amounts to about 19 million unique Tweets for our analysis.

For identifying *fans*, *haters*, and *neutrals*, one could think of simply using the hashtag used by a particular Twitter user. However, such an approach would be misleading since a neutral consumer may write a Tweet about a match using hashtags for either of the teams, while a fan of one team may tweet and mention a hashtag of another team. We propose a more sophisticated way of identifying *fans*, *haters*, and *neutrals* that uses the sentiment expressed in Tweets. In general, 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. In this study, we generate 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 Natural Language Processing. In contrast to neural networks – a high performing algorithm though black box approach – the RF estimator shows how important individual features are in determining outcomes. Our model was trained on more than three million features or word tokens with the then most important features

[5] 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.

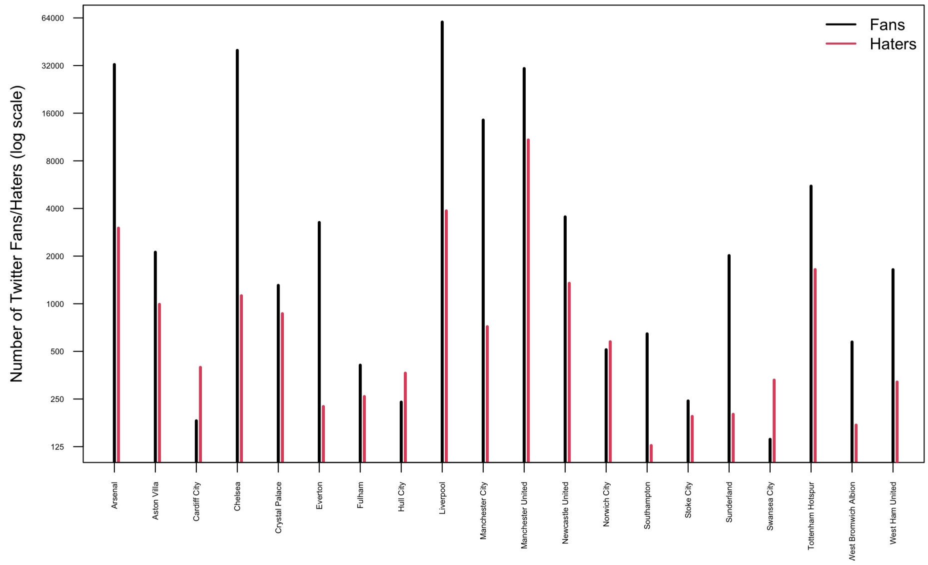
being *bad*, *performance*, *best*, *n't*, *funny*, *dull*, *great*, *like*, *good*, and *waste* (see Appendix A.4 for further details on the architecture of the winning model).

We then 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. If the user does not comment positively on more than two wins 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 look at Tweets with hashtags associated with the losing team, i.e., post-*loss* sentiment. Again, ranked by average sentiment score of tweets, we take the most commonly occurring team in each user's top 5. The inutition 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, hateship cannot be found. The remainder of users are neither denoted as a fan nor as a hater and are assumed to be neutral.[6]

In general, most users tweet about a team post win rather than post loss, with eligible users (users with at least 3 tweets) tweeting 1,221,340 times about the winner of a team post-win and only 629,637 times about the loser of a match post-loss. Based on our rule-based approach, the user's top scoring is heavily favored in determining fandom vs hatership. Post win, the average top scoring sentiment is 15.09, well above the average sentiment score overall of 13.43. Post loss, however, the average top scoring sentiment is 13.86 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.

[6] If we find overlap between fandom and hateship of the *same* team for a Twitter user (many users regularly comment on just one or two clubs), we assign either fandom or hateship according to the higher absolute value of the post-match sentiment score. 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.

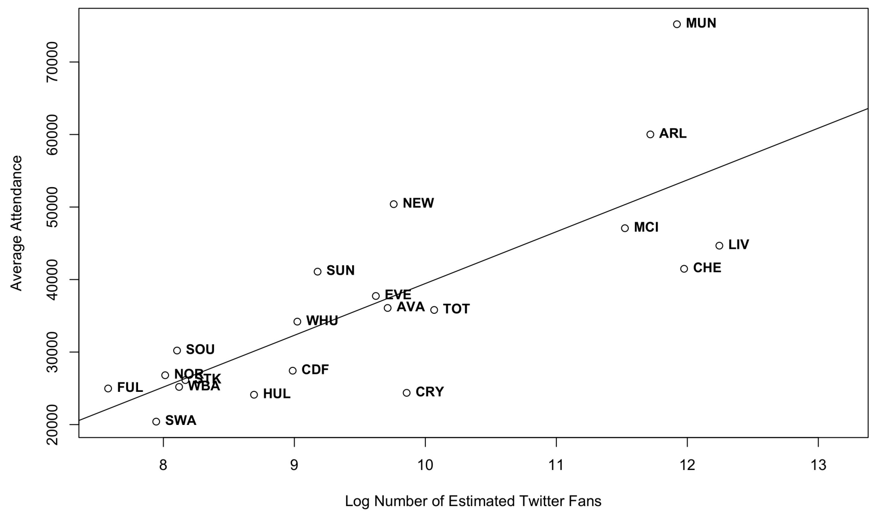
Figure 1: Number of Twitter users regarded as *fans* and *haters* of a particular team.



Following this approach, we identified 196,270 users as *fans*, 23,747 users as *haters*, 3,792 users as both a *fan* of one team and a *hater* of another team, and 1,096,225 users as *neutrals* amongst the overall 1.3 million Twitter users. Figure 1 provides an overview on the number of Twitter users regarded as *fans* and *haters* of particular teams.

In order to see how 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 considerably reduced chance of winning the title. Out of overall 214,133 'Liverpool' Tweets before, during, and after this match, 28 percent are by Liverpool fans as identified by our approach. As expected, of the 5,577 users retweeting "@LFC LOL!" after the game, only a marginal portion of these users (2.67%) are Liverpool fans as identified by our approach. 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 Figure 2) as well as the number of (actual) followers of the official team

Figure 2: Fans identified from sentiment and average attendance.



Notes: This Figure plots the logarithmized number of identified fans following the method as described in Section 2 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.

accounts (see Figure 3) adding some further credibility to our approach.[7]

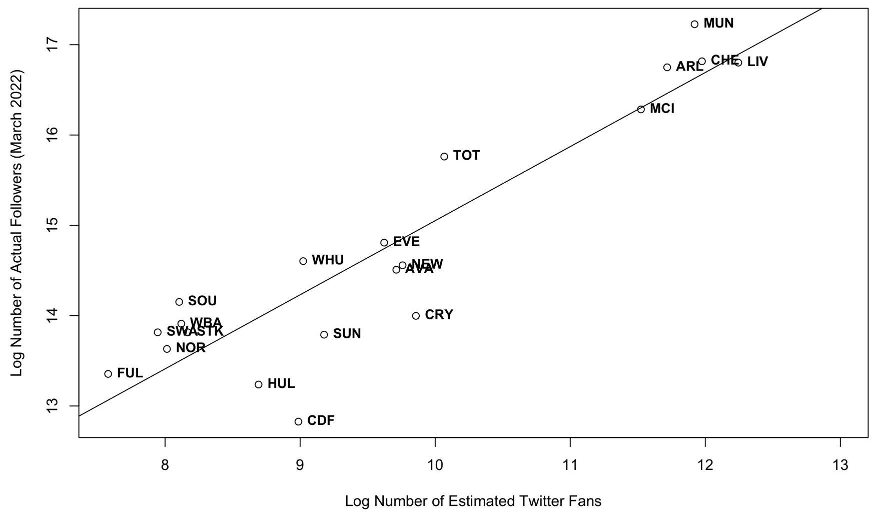
2.2 Measuring Emotional Cues

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 measureing 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 preapred to both buy and sell betting contracts. However, while some studies have shown that *Betfair*, or betting exchange, prices, accurately predict outcomes (Croxson and Reade, 2014), others have rejected the hypothesis of semi-strong market efficiency.

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

Figure 3: Fans identified from sentiment and followers of official team accounts.



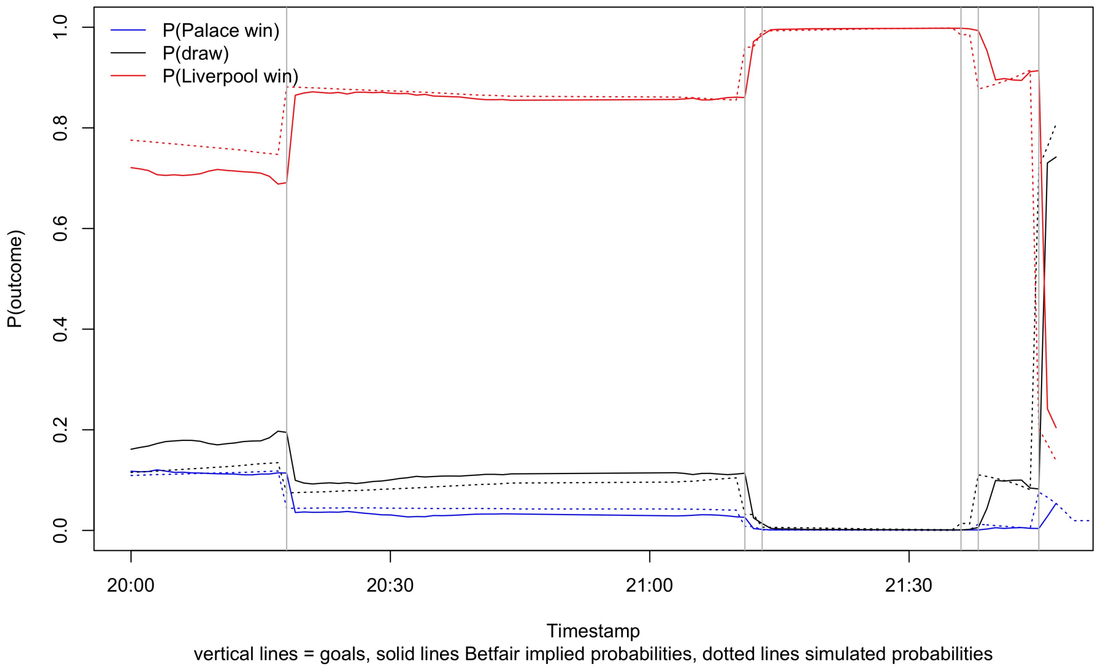
Notes: This Figure plots the logarithmized number of identified fans following the method as described in Section 2 and the logarithmized number of followers of team accounts as of March 2022. ARL: Arsenal, AVA: Aston Villa, CDF: Cardiff City, CHE: Chelsea, CRY: Crystal Palace, EVE: Everton, FUL: Fulham, 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.

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).

In order to see 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. Figure 4 shows how actual and simulated probabilities developed during the course of this match.

Figure 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').

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:

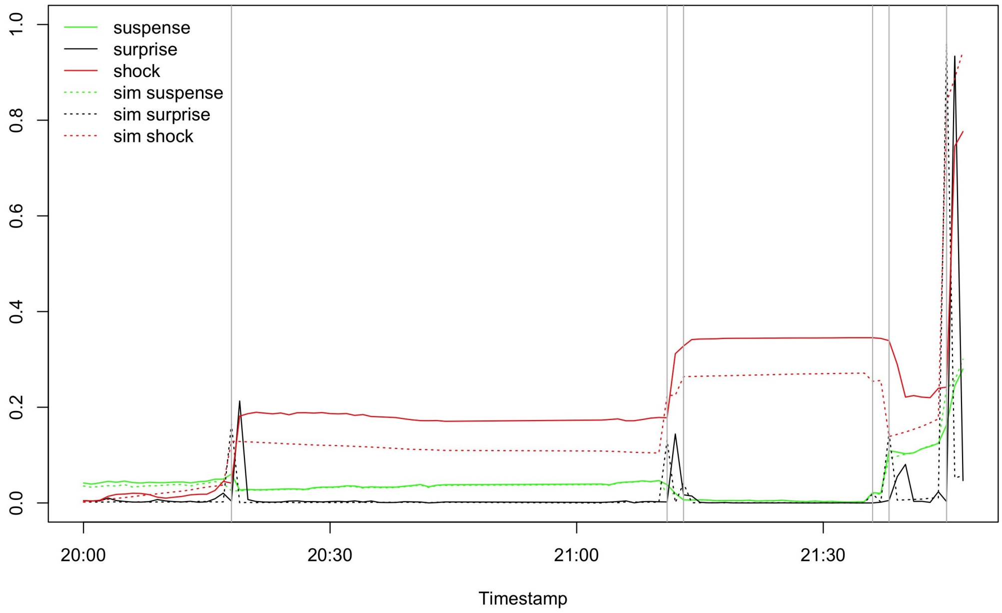
$$Shock_t = \sqrt{(p_t^H - p_0^H)^2 + (p_t^D - p_0^D)^2 + (p_t^A - p_0^A)^2}. \tag{2}$$

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

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

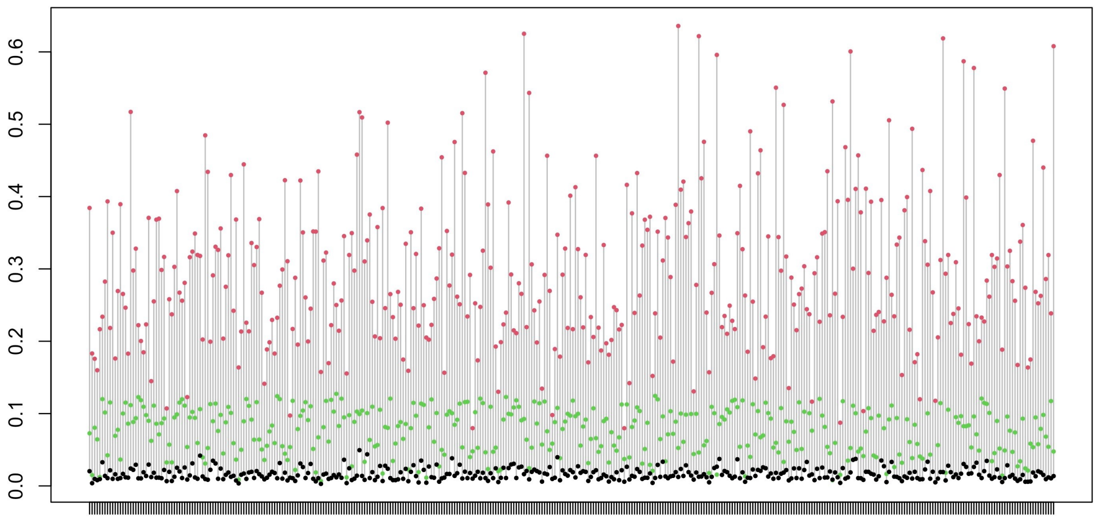
Taking the same example as before, Figure 5 indicates how shock, surprise, and suspense develop during the course of the match. Overall, the observed patterns seem reasonable. While suspense gradually *decreases* up to the 79th minute when Crystal

Figure 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 3 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').

Figure 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.

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, however, the pattern of surprise is spiky and mainly depends on (un-)expected goals scored. Finally, it is worth noting that we not only observe variation in shock, surprise, and suspense *within* a match but also *between* matches (see Figure 6). This must be considered in our empirical model.

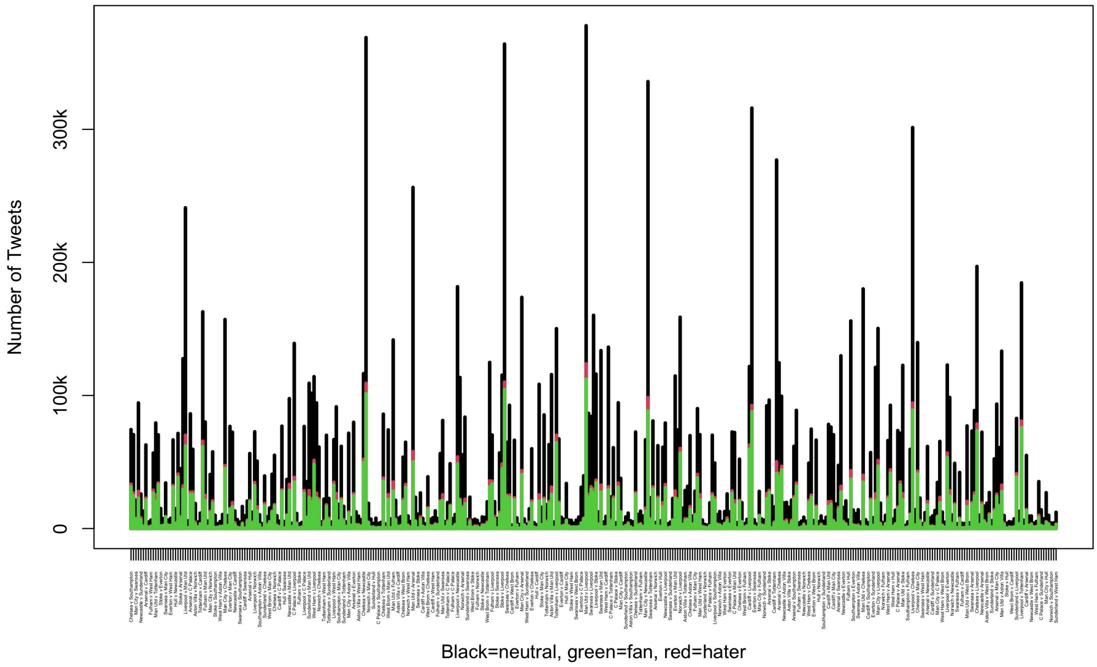
2.3 Empirical Model

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 that include *home* team and/or *away* team hashtags in a given minute t of match i serves as the dependent variable y_{it} in our empirical model:

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 surprise_{it} + \beta_3 shock_{it} \\ + \beta_4 suspense_{it} + \beta_5 X_{it} + \gamma_t + \nu_i + u_{it}. \end{aligned} \tag{4}$$

In order to separate the *net* effects of our emotional cue measures, we control for lagged number of Tweets y_{it-1} and a set of in-match events X_{it} like goals scored, shots, corners, cards, or substitutions. Note that as X_{it} includes total goals scored, it could be seen as a kind of basic 'excitement' index. In order to pick up any differences between minutes played and across matches, we control for minute fixed effects γ_t and match fixed effects ν_i. We also run these regressions seperately for Twitter users that we have identified as *fans*, *haters*, and *neutrals*. That is, for a match involving two teams, we count the tweets of fans (haters) of each team separately if they send a tweet using a hashtag for their favoured (hated) team. The count of neutral tweets for a match is made up of both neutral users *and* 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. We observe a remarkable variation in number of Tweets by *fans*, *haters*, and *neutrals* across the matches in our sample; in Figure 7 we plot these three counts for every match in our dataset.

Figure 7: Number of tweets per match and type of Twitter user.



Notes: This Figure plots the number of Tweets per match by neutrals (black), haters (red), and fans (green) using home team hashtags for all 380 matches played in season 2013/2014.

3 Results

Table 1 provides an overview of our regression results separated for *fans*, *haters*, and *neutrals*. Since we control for lagged number of Tweets and excluded extra time, these regressions are based on 89 minutes for 380 games. As we make use of both *home* team and *away* team hastags we end up with 67,590 minute-game observations.[8] While all regressions include minute and match fixed effects as well as lagged number of Tweets, only specifications in columns (2), (4), and (6) also include control variables.

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. These findings are robust to the inclusion of the control variables. The only remarkable difference between our specifications with and without control variables is the larger effect size for *suspense* in the *fan* regression with controls. As could be expected, any response to emotional cues is smallest for *neutrals*. Interestingly, however, *haters* respond stronger than *fans* to such cues.

These findings are not driven by using simulated cues. As shown in Table 2, the results look similar when the cues are based on outcome probabilities derived from *Betfair* exchange data sourced via *Fracsoft* instead of simulated bookmaker probabilities. The only difference is the size of the coefficient for surprise which is about 2–3 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 Figures 4 and 5).[9]

Table 3 displays the results from our main specification using Poisson regressions instead of OLS. While our main findings seem robust regading the choice of the estimator used, we find that surprise is larger for *haters* than *fans* only when including controls.

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

[8] Note, we miss 50 minute-observations. As such, we end up with 67,590 instead of 67,640 minute-game observations (i.e., 89 minutes x 380 games x 2 hashtag types).

[9] 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.

Table 1: Results for tweets by *fans*, *haters*, and *neutrals*

	Dependent variable: log number of Tweets by...					
	(1)	(2)	(3)	(4)	(5)	(6)
	Fans	*Fans*	*Haters*	*Haters*	*Neutrals*	*Neutrals*
Lagged number of Tweets	0.425^{***}	0.423^{***}	0.334^{***}	0.334^{***}	0.471^{***}	0.471^{***}
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Surprise	2.101^{***}	1.776^{***}	2.163^{***}	2.303^{***}	0.703^{***}	0.559^{***}
	(0.381)	(0.665)	(0.501)	(0.876)	(0.114)	(0.199)
Shock	1.572^{***}	1.709^{***}	3.515^{***}	3.471^{***}	0.882^{***}	0.875^{***}
	(0.177)	(0.253)	(0.233)	(0.233)	(0.053)	(0.053)
Suspense	-3.883^{***}	-6.469^{***}	-7.527^{***}	-6.591^{***}	-1.851^{***}	-1.689^{***}
	(0.752)	(0.814)	(0.989)	(1.071)	(0.225)	(0.244)
Minute FEs	✓	✓	✓	✓	✓	✓
Match FEs	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓
Observations	67,590	67,590	67,590	67,590	67,590	67,590
R^2	0.447	0.448	0.343	0.344	0.784	0.784
Adjusted R^2	0.443	0.444	0.339	0.339	0.782	0.783
Residual Std. Error	5.192	5.187	6.830	6.829	1.555	1.555
	(df = 67,117)	(df = 67,109)	(df = 67,117)	(df = 67,109)	(df = 67,117)	(df = 67,109)

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 (as indicated by hashtags associated with the corresponding *home* team or *away* team) by *fans* (Columns 1 and 2), *haters* (Columns 3 and 4), and *neutrals* (Columns 5 and 6) serves as dependent variable in the models. All models include minute and match fixed effects. Specifications in Columns (2), (4), and (6) also include control variabels, i.e., dummy variables indicating goal, shot, shot hit goalframe, corner, yellow card, red card, or substitution, as well as total goals scored. Significance levels are *p<0.1; **p<0.05; ***p<0.01.

Table 2: Results for tweets by *fans*, *haters*, and *neutrals* using bookmaker brobabilities

	Dependent variable: log number of Tweets by...					
	(1)	(2)	(3)	(4)	(5)	(6)
	Fans		Haters		Neutrals	
Lagged number of Tweets	0.423***	0.420***	0.331***	0.331***	0.467***	0.467***
	(0.016)	(0.016)	(0.012)	(0.012)	(0.057)	(0.057)
Surprise	5.462***	5.703***	9.179***	8.931***	2.562***	2.511***
	(0.486)	(0.489)	(0.849)	(0.847)	(0.209)	(0.201)
Shock	1.589***	1.744***	3.532***	3.544***	0.940***	0.952***
	(0.193)	(0.211)	(0.290)	(0.288)	(0.136)	(0.135)
Suspense	−3.168***	−6.300***	−6.287***	−6.132***	−1.736***	−1.886***
	(0.707)	(0.976)	(1.014)	(1.159)	(0.380)	(0.414)
Minute FEs	✓	✓	✓	✓	✓	✓
Match FEs	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓
Observations	67,590	67,590	67,590	67,590	67,590	67,590
R^2	0.449	0.451	0.347	0.347	0.786	0.786
Adjusted R^2	0.445	0.447	0.342	0.343	0.784	0.785
Residual Std. Error	5.184	5.177	6.811	6.810	1.548	1.548
	(df = 67,117)	(df = 67,109)	(df = 67,117)	(df = 67,109)	(df = 67,117)	(df = 67,109)

Notes: This Table provides an overview of the effects of emotional cues on the number of Tweets across Twitter users. In contrast to Table 1, all cues are based on outcome probabilities derived from *Betfair* exchange data sourced via *Fracsoft* instead of simulated bookmaker probabilities. The logarithmized number of Tweets (as indicated by hashtags associated with the corresponding *home* team or *away* team) by *fans* (Columns 1 and 2), *haters* (Columns 3 and 4), and *neutrals* (Columns 5 and 6) serves as dependent variable in the models. All models include minute and match fixed effects. Specifications in Columns (2), (4), and (6) also include control variabels, i.e., dummy variables indicating goal, shot, shot hit goalframe, corner, yellow card, red card, or substitution, as well as total goals scored. Significance levels are *p<0.1; **p<0.05; ***p<0.01.

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 equaliser, that effect is covered in the normal *surprise* coefficient.

From our results in Table 4, the main findings remain, i.e., the effects of *surprise* and *shock* are positive while the effects of *suspense* are negative. *Suspense*, however, is not a precise predictor of Twitter activity anymore. Moreover, the interaction effects between winning/losing as well as *surprise* and *shock* are either negative or non-significant suggesting that *surprise* and *shock* unfold their largest effects when the match is currently

Table 3: Results for tweets by *fans*, *haters*, and *neutrals* (Poisson regressions)

	Dependent variable: number of Tweets by...					
	(1)	(2)	(3)	(4)	(5)	(6)
	Fans		Haters		Neutrals	
Lagged number of Tweets	0.777^{***}	0.777^{***}	0.161^{***}	0.160^{***}	0.833^{***}	0.834^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)
Surprise	0.869^{***}	0.212^{***}	0.566^{***}	0.373^{***}	0.650^{***}	0.298^{***}
	(0.010)	(0.016)	(0.033)	(0.052)	(0.007)	(0.011)
Shock	0.293^{***}	0.280^{***}	0.871^{***}	0.714^{***}	0.271^{***}	0.263^{***}
	(0.004)	(0.004)	(0.011)	(0.013)	(0.002)	(0.003)
Suspense	-1.651^{***}	-1.416^{***}	-3.123^{***}	-2.930^{***}	-0.909^{***}	-0.785^{***}
	(0.022)	(0.022)	(0.066)	(0.067)	(0.013)	(0.014)
Minute FEs	✓	✓	✓	✓	✓	✓
Match FEs	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓
Observations	67,590	67,590	67,590	67,590	67,590	67,590
McFadden's Pseudo-R^2	0.839	0.84	0.521	0.521	0.874	0.875
	(df = 67,117)	(df = 67,109)	(df = 67,117)	(df = 67,109)	(df = 67,117)	(df = 67,109)

Notes: This Table provides an overview of the effects of emotional cues on the number of Tweets across Twitter users based on Poisson Regressions. The number of Tweets (as indicated by hashtags associated with the corresponding *home* team or *away* team) by *fans* (Columns 1 and 2), *haters* (Columns 3 and 4), and *neutrals* (Columns 5 and 6) serves as dependent variable in the models. All models include minute and match fixed effects. Specifications in Columns (2), (4), and (6) also include control variabels, i.e., dummy variables indicating goal, shot, shot hit goalframe, corner, yellow card, red card, or substitution, as well as total goals scored. Significance levels are *p<0.1; **p<0.05; ***p<0.01.

a tie. Importantly, following the earlier interpretation of our approach, this holds true even in situation when a goal is scored. In other words, our findings suggest that *goal-induced* effects from *surprise* and *shock* on Twitter activity are the largest, when the favorite (or hated) team either scores or concedes an equaliser. Finally, very suspenseful moments during a match when 'their' team is losing seem to increase the number of Tweets by *fans* but not by *haters* while the corresponding effects from *suspense* remain negative for both *fans* and *haters* when 'their' team is winning.

Table 4: Results for tweets by *fans* and *haters* considering winning and losing

	Dependent variable: log number of Tweets by...			
	(1)	(2)	(3)	(4)
	Fans	*Fans*	*Haters*	*Haters*
Lagged number of Tweets	0.395***	0.395***	0.321***	0.320***
	(0.015)	(0.015)	(0.011)	(0.011)
Surprise	3.686***	3.170***	3.858***	3.812***
	(0.722)	(0.844)	(1.077)	(1.191)
Shock	2.569***	2.748***	3.939**	4.678**
	(0.966)	(0.977)	(1.920)	(1.978)
Suspense	−6.868*	−6.959*	−6.335	−7.365
	(3.884)	(3.894)	(8.427)	(8.660)
Winning	2.475***	2.454***	2.858***	2.798***
	(0.253)	(0.253)	(0.423)	(0.431)
Losing	−1.433***	−1.536***	−0.528	−0.889**
	(0.246)	(0.253)	(0.413)	(0.427)
Surprise x winning	−2.096**	−2.047**	−1.485	−1.019
	(0.942)	(0.916)	(1.483)	(1.463)
Surprise x losing	−2.786***	−2.727***	−4.031***	−3.532***
	(0.885)	(0.863)	(1.316)	(1.287)
Shock x winning	−4.470***	−4.633***	−5.606***	−6.297***
	(1.134)	(1.137)	(2.083)	(2.125)
Shock x losing	0.786	0.537	1.952	0.956
	(1.136)	(1.142)	(2.081)	(2.134)
Suspense x winning	0.413	1.012	−0.051	2.370
	(4.804)	(4.808)	(8.832)	(9.020)
Suspense x losing	8.014*	9.323**	−0.747	4.246
	(4.163)	(4.187)	(8.841)	(9.008)
Minute FEs	✓	✓	✓	✓
Match FEs	✓	✓	✓	✓
Controls		✓		✓
Observations	67,590	67,590	67,590	67,590
R^2	0.459	0.459	0.350	0.350
Adjusted R^2	0.455	0.455	0.345	0.346
Residual Std. Error	5.138 (df = 67,109)	5.137 (df = 67,101)	6.797 (df = 67,109)	6.794 (df = 67,101)

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 (as indicated by hashtags associated with the corresponding *home* team or *away* team) by *fans* (Columns 1 and 2) and *haters* (Columns 3 and 4) serves as dependent variable in the models. All models include minute and match fixed effects. Specifications in Columns (2) and (4) also include control variabels, i.e., dummy variables indicating goal, shot, shot hit goalframe, corner, yellow card, red card, or substitution, as well as total goals scored. In contrast to results presented in Table 1, all specification also include variables which measure whether the favorite (or hated) team is currently winning or losing as well as the corresponding interactions with all emotional cues. Significance levels are *p<0.1; **p<0.05; ***p<0.01.

4 Discussion and Conclusions

By analyzing 19m tweets in combination with in-play information for overall 380 games played in the English Premier League we provide empirical evidence that emotional cues significantly influence Twitter activity. 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* - on average - decreases the number of Tweets. As could be expected, any response to emotional cues is smallest for *neutrals*. Interestingly, however, *haters* respond stronger than *fans* to such cues. Further analysis suggests that *goal-induced* effects from *surprise* and *shock* on Twitter activity are the largest, when the favorite (or hated) team either scores or concedes an equaliser. Moreover, we observe some asymmetries regarding the response to *suspense*. Very suspenseful moments during a match when 'their' team is losing increase the number of Tweets by *fans* but not by *haters*. At the same time, however, the corresponding effects from *suspense* remain negative for both *fans* and *haters* when 'their' team is winning.

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.

Overall, 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 influcene immediate human behavior. Our setting seems 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 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 offer a very fine-grained empirical test for the uncertainty-of-outcome hypothesis in sports. In fact, we find 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). We argue that 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)).

References

Angelini, G., De Angelis, L., and Singleton, C. (2022). Informational efficiency and behaviour within in-play prediction markets. *International Journal of Forecasting*, 38(1):282–299.

Babad, E. and Katz, Y. (1991). Wishful thinking—against all odds. *Journal of Applied Social Psychology*, 21(23):1921–1938.

Bizzozero, P., Flepp, R., and Franck, E. (2016). The importance of suspense and surprise in entertainment demand: Evidence from Wimbledon. *Journal of Economic Behavior & Organization*, 130:47–63.

Buraimo, B., Forrest, D., McHale, I., and Tena, J. (2020). Unscripted drama: Soccer audience response to suspense, surprise, and shock. *Economic Inquiry*, 58(2):881–896.

Caplin, A. and Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116(1):55–79.

Choi, D. and Hui, S. (2014). The role of surprise: Understanding overreaction and underreaction to unanticipated events using in-play soccer betting market. *Journal of Economic Behavior & Organization*, 107:614–629.

Croxson, K. and Reade, J. (2014). Information and Efficiency: Goal Arrival in Soccer Betting. *Economic Journal*, 124:62–91.

Dillenberger, D. (2010). Preferences for one-shot resolution of uncertainty and Allais-type behavior. *Econometrica*, 78(6):1973–2004.

Ely, J., Frankel, A., and Kamenica, E. (2015). Suspense and surprise. *Journal of Political Economy*, 123(1):215–260.

Fischer, L., Kelava, A., Nagel, M., and Pawlowski, T. (2023). Celebration beats frustration – emotional cues and alcohol use during soccer matches. *University of Tübingen: mimeo.*

Gul, F. (1991). A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society*, pages 667–686.

Jiwa, M., Cooper, P., Chong, T. T.-J., and Bode, S. (2021). Choosing increases the value of non-instrumental information. *Scientific Reports*, 11(8780).

Kaplan, S. (2021). Entertainment utility from skill and thrill. *SSRN Working Paper*, doi.org/10.2139/ssrn.3888785.

Kőszegi, B. and Rabin, M. (2009). Reference-dependent consumption plans. *American Economic Review*, 99(3):909–36.

Kreps, D. and Porteus, E. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, pages 185–200.

Liu, X., Shum, M., and Uetake, K. (2021). Passive vs. active attention to baseball telecasts: implications for content (re-)design. *SSRN Working Paper*, doi.org/10.2139/ssrn.3717894.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2):426–432.

Lucas, G. M., Gratch, J., Malandrakis, N., Szablowski, E., Fessler, E., and Nichols, J. (2017). GOAALLL: Using sentiment in the world cup to explore theories of emotion. *Image and Vision Computing*, 65:58–65.

Neale, W. C. (1964). The Peculiar Economics of Professional Sports. *Quarterly Journal of Economics*, 78(1):1–14.

Palacios-Huerta, I. (1999). The aversion to the sequential resolution of uncertainty. *Journal of Risk and Uncertainty*, 18(3):249–269.

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

Raney, A. A. (2018). Why we watch and enjoy mediated sports. In Raney, A. A. and Bryant, J., editors, *Handbook of sports and media*, pages 313–329. Lawrence Erlbaum Associates, New York, NY.

Richardson, T., Nalbantis, G., and Pawlowski, T. (2023). Emotional cues and the demand for televised sports: Evidence from the UEFA Champions League. *Journal of Sports Economics*, forthcoming.

Rottenberg, S. (1956). The Baseball Players' Labor Market. *The Journal of Political Economy*, 64(3):242–258.

Simonov, A., Ursu, R. M., and Zheng, C. (2022). EXPRESS: Suspense and surprise in media product design: Evidence from Twitch.tv. *Journal of Marketing Research*, doi.org/10.1177/00222437221108653.

Zillmann, D. and Cantor, J. R. (1972). Directionality of transitory dominance as a communication variable affecting humor appreciation. *Journal of Personality and Social Psychology*, 24(2):191–198.

Appendix

For our research purposes, we use in-game *Betfair* exchange data sourced via *Fracsoft*. *Betfair*, a peer-to-peer platform, is the largest online betting exchange in the world. Unlike alternative bookmakers, *Betfair* prices (odds) are readily available via the *Betfair* Application Programming Interface (API). Specifically, we work with csv files, compiled by *Fracsoft*, which include match descriptions, scheduled and actual game times, timestamps (UTC) for each price movement, an in-play dummy variable, betting market status (open or closed), and volumes and odds available for each selection (home, away, and draw). Focusing on the 2013/2014 season of the English Premiere League, we collect data for all 380 matches between the 20 EPL teams. With price movements and betting volumes featuring granularity to the millisecond, we use in-match odds to impute the real-time outcome probabilities that are eventually used to measure our emotional cues – shock, surprise, suspense – through the course of a match. Preceding the calculation of these cues, the following procedures are performed using python scripts toward the Fracsoft dataset, aggregated by the minute.

A.1 Prices and Actual Probabilities

For any given minute within a match there may be more than one unique exercised price match. In this case, a pre-determined aggregation function must be used to determine the value we use. We employ three different methods. Mean, the strategy we default to for our calculations, is simply the mean of every unique price match found in the minute. Similarly, we explore the median of these set of price matches by the minute. Finally, we consider a scheme weighted by volume that we've named "effective odds". Alternative to the other methods, the "effective odds" are proportional to the volumes of each price match. For example, suppose we find that a given minute has price matches for 1.5 and 2. Respectively, the volumes traded are 10 and 90. Using the mean method we would find the aggregated price match to be 1.75 regardless of volume. In contrast, we use volume proportional pricing to calculate 1.95 as our effective odds. Intuitively, these odds tend toward the price match with the higher amount of volume. In practice, however, we found little to no difference in our subsequent results. As a result, we opt to present mean, the more straight-forward aggregation method.

Once the odds are appropriately aggregated, we then derive implied outcome probabilities by taking the inverse odds. In theory, the sum of the inverse odds for every possible outcome should be equal to 1. In reality, however, the sum of these values is slightly above 1. The difference between the sum of implied outcome probabilities and 1 is known as the overround or vig – essentially the bookmakers fee or commission. In order to remove this overround, we proportionally scale the derived outcome probabilities. More specifically, we divide each outcome probability by the sum of all outcome probabilities.

A.2 Simulated Probabilities

While some studies have shown that *Betfair*, or betting 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. As such, we use in-play odds derived from an in-play model as proposed by Buraimo et al. (2020) in our main specification.

Assuming an independently Poisson distributed number of goals scored by home and away, we estimate team-specific scoring rates by minimizing the squared difference between the bookmaker implied probabilities (imputed using odds on over/under totals) and the outcome probabilities from the in-play model. Furthermore, rather than assuming identical distribution of goals between every minute in a match, we use our goals scored data to distribute historical scoring rates across the minutes. Accounting for the average amount of injury time - and to share out the inflated scoring rates found in the 45th and 90th minute - we presume all matches to be 93 minutes long. Then we calculate a moving average over 15 min to smooth the relative frequency distributions. Using backward interpolation, we fill in the missing values for the first 15 minutes. Finally, we calculate the density function of goals scored per minute.

With goal distributions and team-specific scoring rates as our main ingredients, we execute the match simulations that enable us to calculate hypothetical probabilities. For every match in our dataset, we simulate the number of goals in each minute, sum up the score line, and record the result. We repeat this simulation 100,000 times per minute per match (9 million simulations for every match). In pursuance of runtime reduction, we execute concurrent simulations using python's built-in multithreading packages and distributed computation. For any given minute, the respective outcome probabilities are represented by the number of simulations with each outcome, given the current score, divided by the total number of simulations.

A.3 Events

Match events depicted throughout our report were sourced from match commentaries supplied by whoscored.com. Other sources considered included BBC and ESPN, however, whoscored.com proved to have the most extensive database. Exactly 12 event types were collected from the public site; events include goal scored, save made, card received, offside, corner, attempt missed, attempt blocked, woodwork hit, substitution off, substitution on, start half, and end half. Most notably, the events "goal scored" and "red card received" were used during analysis and to generate both in-play models (explained in further depth later) respectively. The predominant tools used to acquire this data were python and selenium, a python package primarily used for test automation. Presented in order, data collection included a complete acquisition of urls corresponding to every unique match found in the Odds dataset. Then, using the browser automation enabled by selenium, we systematically scanned each of the gathered urls for in-game commentaries with timestamps, down to the second, included. In addition to having the most exhaustive catalog, whoscored.com commentaries appear to have the most granular timestamps. This added granularity proved to be crucial in our in-game analysis. Finally, the data acquired was packaged into individual xml files corresponding to each match. Using an element tree structure, every commentary entry is presented as a sub element of the larger commentary tree. Root attributes include away team, home team, season, season id, game date and time, league name, sport name, and language. Sub elements include the attributes comment, period, minute, second, expanded minute, and event type. With the complete timestamps included in these xml files, the data is ultimately merged into the master dataset and adjusted by the historical start times found in the Fracsoft dataset. Since we aggregate events by minute as well, we transform single events into a comma delimited string of events in chronological order. Neutral events such as the start and end of a half are regarded as home team events.

A.4 Sentiment Analysis and Fandom

The random forest - a supervised machine learning model - is an ensemble of decision trees popularly used for sentiment analysis. Although deep learning Neural Nets like the Long Short Term Memory algorithm can hypothetically outperform tree-based models for sentiment analysis, they are “black box” approaches with no discernable feature importances. Alternatively, the random forest shows how important individual features are in determining outcomes. Our machine learning scripts use the python packages nltk and sklearn. Specifically, nltk was used for preprocessing and we employed sklearn for model training and evaluation. Since we're interested in obtaining sentiment magnitudes, we use a regressor rather than a categorical classifier.

Our sentiment analysis model is trained on data from Stanford's Sentiment Tree Bank. After considering multiple open-source sentiment datasets, we found that the Stanford data consisted of more realistic use cases perhaps more relevant to our own twitter data. Before training, we follow traditional NLP prepossessing protocols. We remove English stop words, replace broken conjugations, and remove noise caused by encoding errors. Again, using nltk, we lemmatize in order to find root words, and use sklearns TfidfVectorizer to extract our feature set. We include our entire twitter dataset into our corpus to ensure all features are extracted. Instead of training on all the observations, each tree of RF is trained on a subset of the observations.

After performing an extensive grid search to tune model hyperparameters, our final specification is used to generate sentiment scores, ranging from 0 to 25, for every tweet associated with every match in our dataset. In summary, in our winning model (accuracy: 70.08%, MAE: 2.49, MSE: 13.88), the maximal depth of a tree, which is defined as the longest path between the root node and the leaf node, was 90. We used 'auto' for the number of features to consider when looking for the best split. The minimum number of samples required to be at a leaf node was two, the minimum number of samples required to split an internal node was nine, and the number of trees in the forest was 550.

As described in Section 2, we curate a rule set in order to assign fan association as well as define haters and neutral spectators.

A.5 Shock and Surprise

The two backward looking emotional cues, shock and surprise, are similar in their calculations. Surprise essentially refers to difference in beliefs relative to preceding events, whereas shock captures the contrast between in-game beliefs and the initial projections. Accordingly, the formula for surprise is the square root of the sum of the squared differences in outcome probabilities and outcome probabilities of the previous time period. Moreover, shock is calculated by taking the square root of the sum of the squared differences in outcome probabilities and outcome probabilities pre match. The key lies in the time dimension of the anterior reference point. For surprise, the outcome probabilities are subtracted by the outcome probabilities of the previous time period. On the other hand, the probabilities are subtracted by a static, pre-match observation. For both measures, we do these operations for all 3 of the aforementioned price aggregations.

A.6 Suspense

Suspense, the forward-looking measure, requires more computation than its backward-looking counterparts. In this regard, we can't rely on historical observations to formulate an in-play model; instead, we take a simulation-based approach very similar to the procedures described in section A.2. With these procedures in place, we leverage simulated match outcomes to calculate scenario contingent, hypothetical probabilities. Essentially, we find the probabilities for a home win, draw, or away win if either team scores in the next minute. For every match we iterate through each minute and find the likelihood of home (away) win given a home (away) goal. To do so we simply isolate the simulations with home (away) goals appearing within the next minute and find the respective proportion of home wins, draws, and away wins given the current score line. We then square the difference between these and the given in-play model probabilities for home win, draw, and away win for the minute. Next, we multiply this squared difference by the probability of a home (away) goal in the next minute and sum all of values found for each outcome. Finally, we define the minute's suspense measure by the square root of the sum of these sums.