

# Catching Stray Balls

Football, fandom, and the impact on digital discourse

Mark J. Hill, King's College London

## Research Question & Motivation

Do real-world events trigger toxic behaviour that spreads across unrelated online communities?

**Toxicity** has been defined as 'interactions directed at an entity designed to be inflammatory' (Hanscom et al. 2024).

### Why it Matters:

- Users encountering toxic online discourse experience negative psychological consequences (Braghieri et al. 2022; Allcott et al. 2020)
- Social media amplifies emotional content through platform design (Milli et al. 2025; Kramer et al. 2014)
- Those exposed to emotionally charged content are more likely to express similar sentiments (Ferrara and Yang 2015; Brady et al. 2017)

**Gap:** Limited understanding of how toxicity emerges and moves *between* communities.

**Approach:** Football as a 'natural experiment' with clear, time-stamped emotional triggers to trace sentiment spillover across digital spaces.

## Data & Methods

**Scale:** 62+ million Reddit posts from 41 football club subreddits (2008-2024)

**Matched Events:** 20,764 match results aligned with posting times and football club subreddit

**Cross-Community Analysis:** Over half a million pairs of posts by same users in football and non-football subreddits (within 10-minute windows)

| Match Result | Posts     | %    |
|--------------|-----------|------|
| Wins         | 6,477,964 | 49.6 |
| Draws        | 2,690,511 | 20.6 |
| Losses       | 3,902,686 | 29.9 |

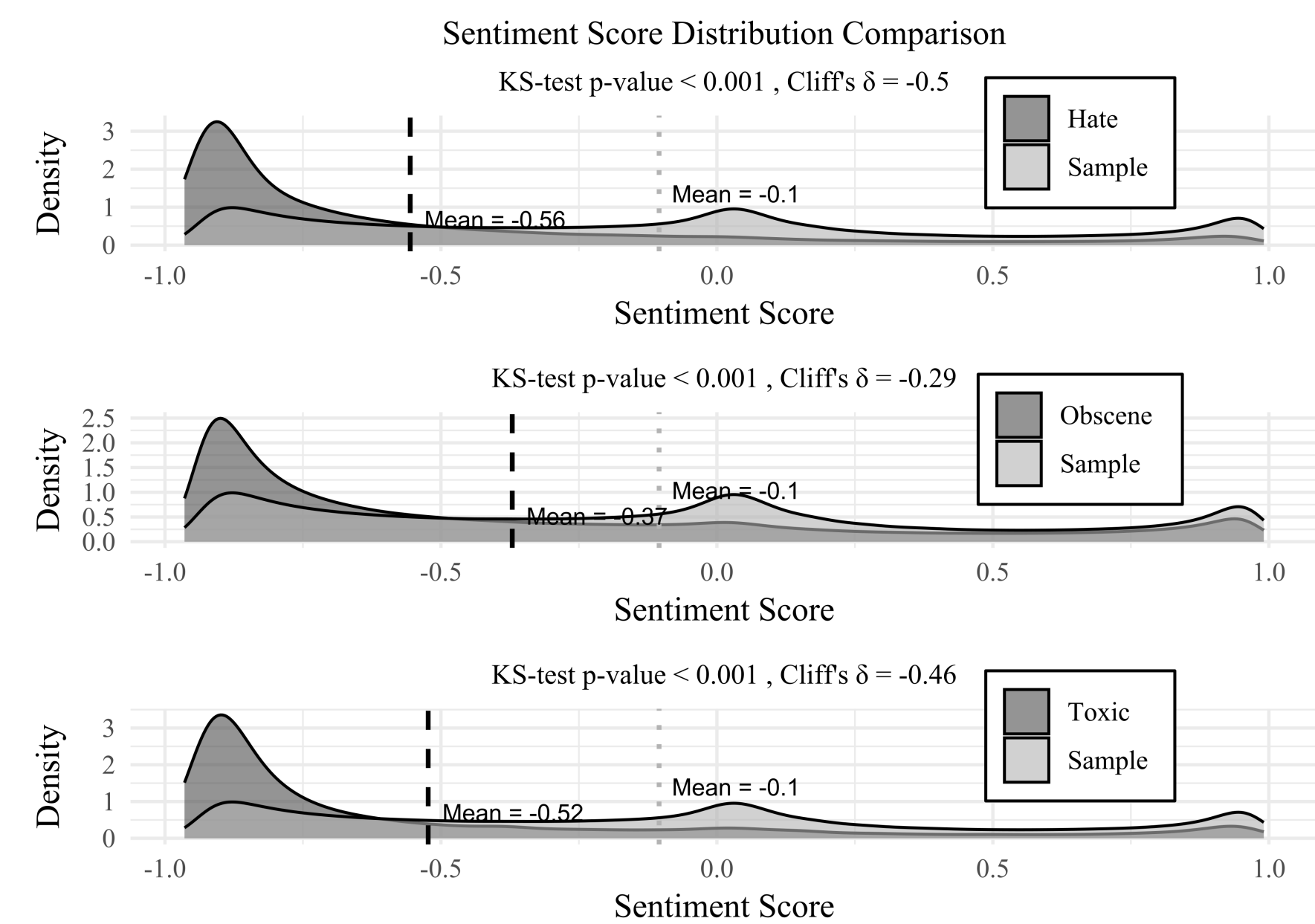
### Analysis Pipeline:

- RoBERTa-based sentiment and toxicity analysis
- Match outcome correlation with sentiment shifts
- Cross-community sentiment spillover measurement
- Linguistic feature analysis of negative content

## Finding 1: Negative Sentiment & Toxicity

### How do we measure toxicity?

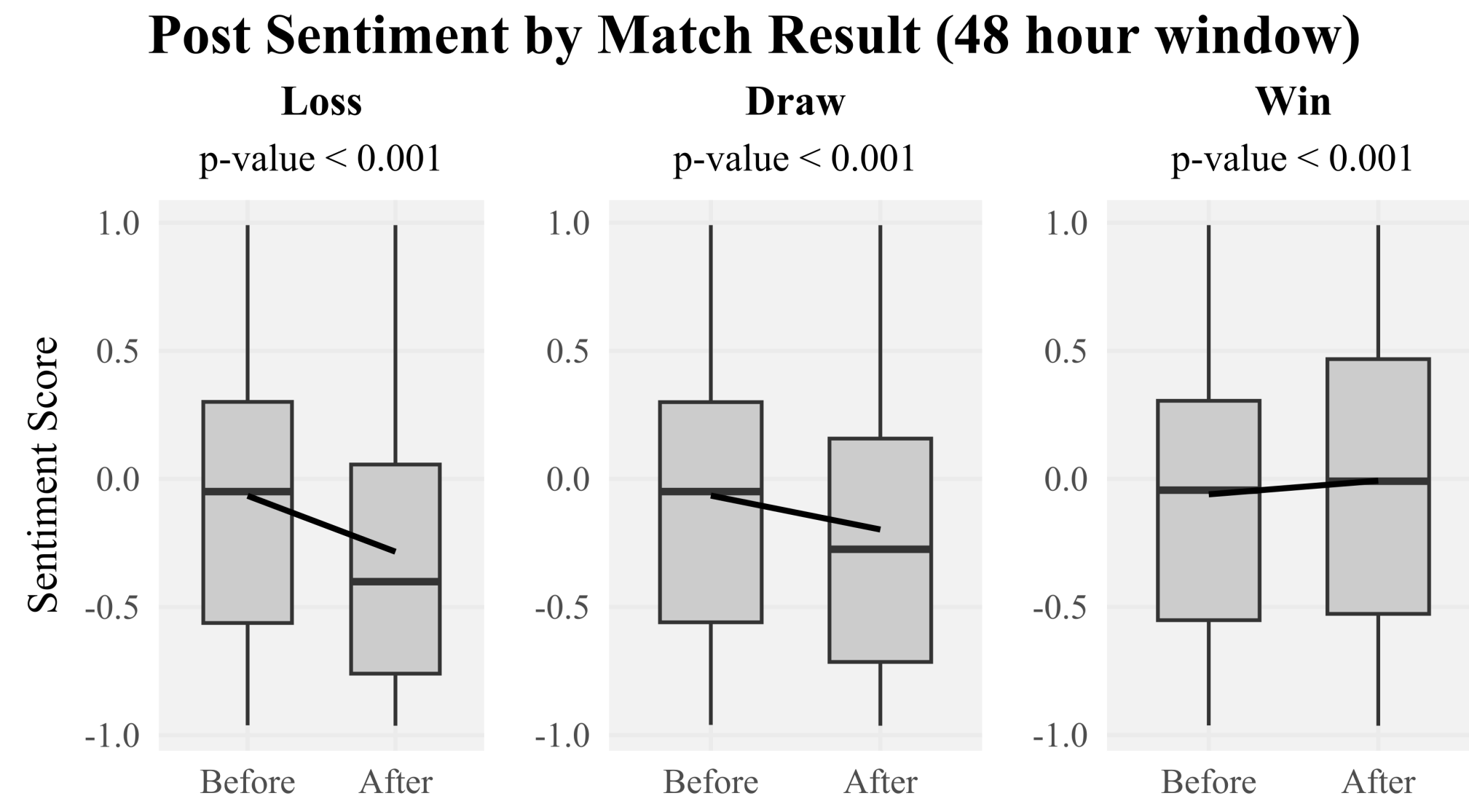
Toxicity is contextual. Posts in football subreddit after a loss may be perceived as negative rather than toxic. That same emotion, however, may be toxic in unrelated subreddits. Therefore we use sentiment and toxicity detection, as well as profanity lexicons to assess the relationship.



**Strong correlation** between negative sentiment and problematic content.

**Implication:** Sentiment is reliable proxy and early warning signal

## Finding 2: Football Results Correlate to Online Sentiment

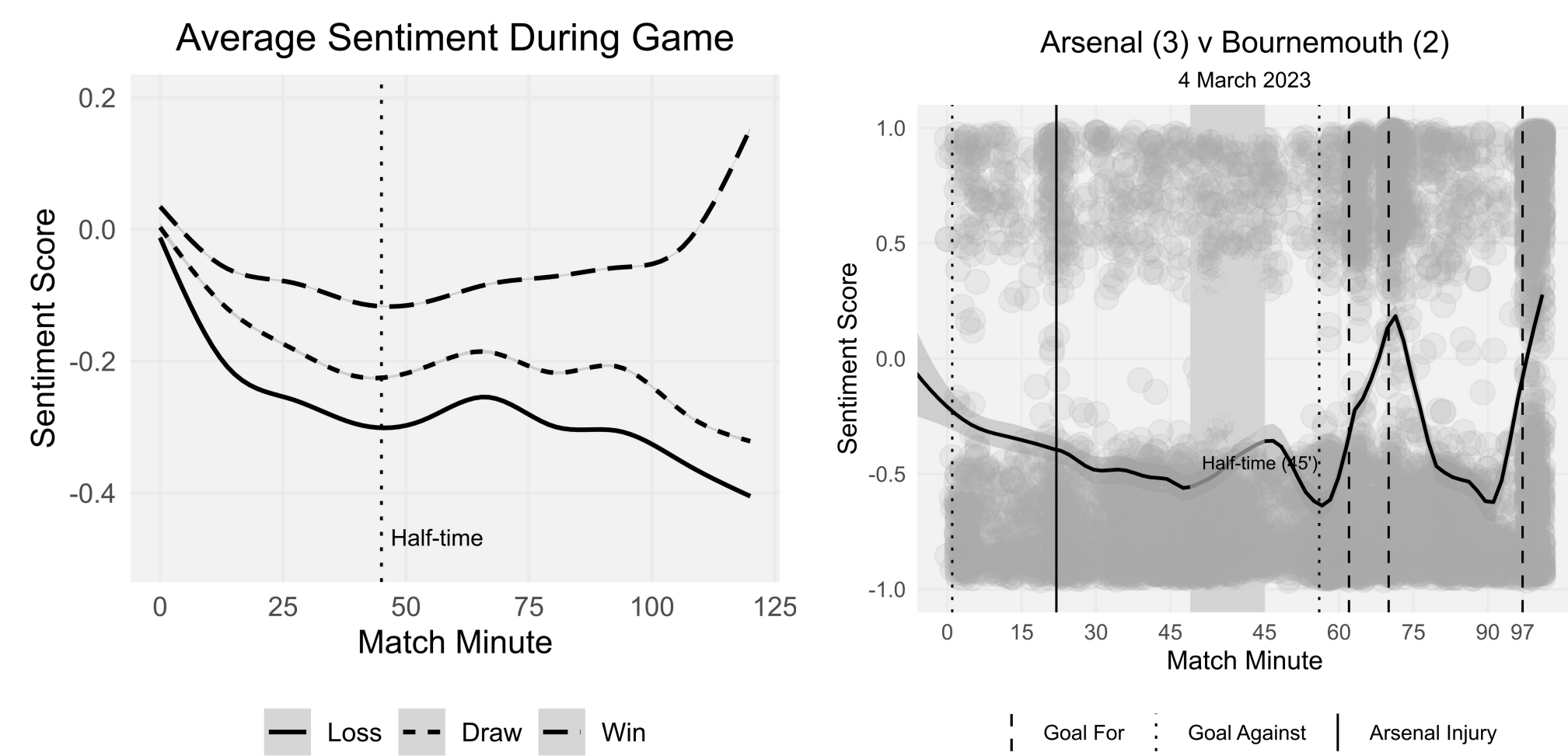


| Result                    | Posts per match | Post ratio | Average Sentiment |
|---------------------------|-----------------|------------|-------------------|
| <i>Within 120 Minutes</i> |                 |            |                   |
| Loss                      | 739             | 0.89***    | -0.25***          |
| Draw                      | 764             | 0.92***    | -0.11***          |
| Win                       | 934             | 1.12***    | 0.07***           |
| <i>Within 8 Hours</i>     |                 |            |                   |
| Loss                      | 577             | 0.80***    | -0.17***          |
| Draw                      | 605             | 0.84***    | -0.09***          |
| Win                       | 904             | 1.26***    | 0.06***           |

**Asymmetric Effect:** Losses decrease sentiment and posting; wins increase posts but have smaller impact on sentiment

**Temporal Correlation:** Clear relationship between match events and sentiment shifts

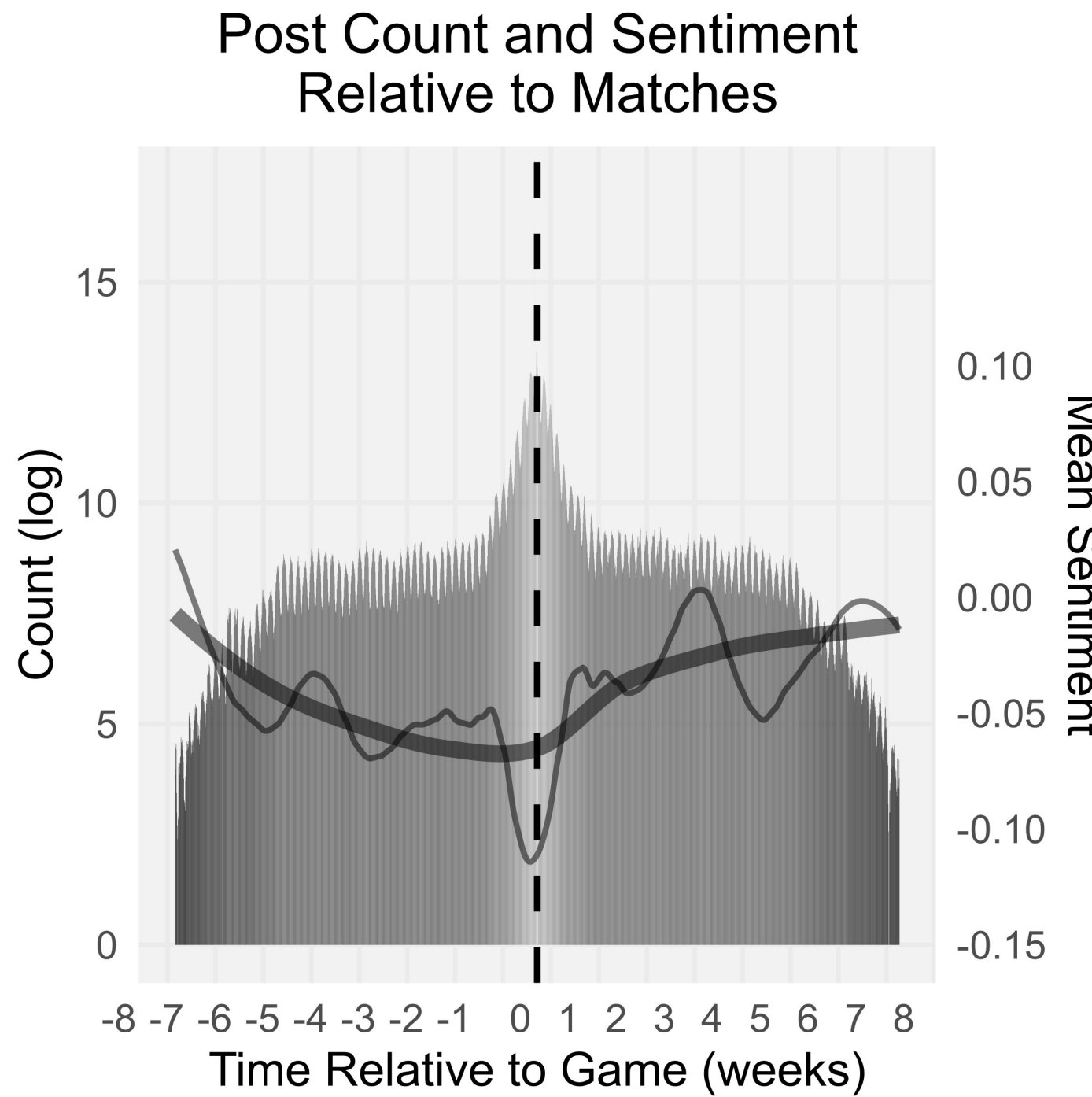
## Real-Time Emotional Dynamics



**Left:** Aggregated sentiment per-minute across all matches

**Right:** Individual match (Arsenal vs Bournemouth, 2023)

**Below:** Post count and sentiment relative to match timing



### Key patterns:

- Universal sentiment drop at kick-off (anxiety)
- Diverging patterns based on (final) match outcome
- Negative reactions more sustained than positive
- Immediate measurable responses to key match events
- Sentiment is lower nearer to matches
- Sentiment decay over time following matches

## Finding 3: Emotional Spillover Across Communities

Emotional states in football subreddits correlate to sentiment in unrelated subreddits

| Time Period      | Kendall's $\tau$ | n       |
|------------------|------------------|---------|
| All Paired Posts | 0.085***         | 575,863 |
| During Matches   | 0.118***         | 234,024 |
| Outside Matches  | 0.059***         | 341,839 |

**Finding:** Small correlation, but relative changes and statistical significance across large sample represent meaningful patterns of user behaviour at scale. The doubling of  $\tau$  during matches suggests real-world events amplify cross-community emotional spillover. This represents (within this sample) thousands of additional negative posts in unrelated subreddits.

|                 | Negative | Neutral | Positive |
|-----------------|----------|---------|----------|
| <b>Negative</b> | 40.00    | -18.76  | -26.73   |
| <b>Neutral</b>  | -18.92   | 22.74   | -7.59    |
| <b>Positive</b> | -27.49   | -5.12   | 44.66    |

$\chi^2$  standardised residuals

**Finding:** Matching sentiment across communities occurs more than chance

## Finding 4: Linguistic Spillover of Toxic Features

**Method:** Measured correlations between linguistic features in paired posts across subreddits.

| Feature       | Outside Match | During Match | Difference ( $\Delta\tau$ ) |
|---------------|---------------|--------------|-----------------------------|
| Profanity     | 0.061***      | 0.109***     | 0.048***                    |
| Violent words | 0.022***      | 0.049***     | 0.027***                    |
| Intensifiers  | 0.059***      | 0.074***     | 0.015***                    |
| Exclamations  | 0.124***      | 0.154***     | 0.035***                    |
| All-caps      | 0.052***      | 0.133***     | 0.081***                    |

**Finding:** All linguistic toxicity markers strengthen during matches

**Implication:** Heightened emotional states from football events intensify toxic communication patterns across unrelated digital spaces

## Implications & Applications

Digital communities are interconnected emotional ecosystems

### For Research:

- Methods to measure emotional contagion cascading
- Applicable beyond football (elections, breaking news)

### For Platform Design:

- Predictive moderation during high-risk events
- Early warnings through cross-platform monitoring
- Indicators for temporarily limiting cross-community posts

### For Society:

- Understanding offline-to-online harm pathways
- Hidden mechanisms of toxicity propagation

**Broader Impact:** Computational evidence of real-world event-driven emotional spillover across unrelated digital communities

**ACL 2025 VIENNA**  
JULY 27 - AUGUST 1

**KING'S**  
College  
LONDON

**Contact:** mark.j.hill@kcl.ac.uk

**Pre-print:** doi.org/10.48550/arXiv.2506.01642

**Code:** github.com/markjhil/2025-catching-strays