Reference: Fletcher, T., Kearns, C., Kilvington, D., Sinclair, G., Lynn, T., Black, J., Doidge, M., Leoni Santos, G., Liston, K., Rodrigo-Jesue, I., & Rosati, P. (2024) Social media and online hate in sport: A case study of association football. In A. Billings., & M. Hardin (eds), *Routledge handbook of sport and social media.* Routledge.

Social media and online hate in sport: A case study of association football

Thomas Fletcher, Leeds Beckett University, Leeds, UK

Colm Kearns, Dublin City University, Dublin, Ireland

Daniel Kilvington, Leeds Beckett University, Leeds, UK

Gary Sinclair, Dublin City University, Dublin, Ireland

Theo Lynn, Dublin City University, Dublin, Ireland

Jack Black, Sheffield Hallam University, Sheffield, UK

Mark Doidge, Loughborough University, Loughborough, UK

Guto Leoni Santos, Dublin City University, Dublin, Ireland

Katie Liston, Ulster University, Newtownabbey, UK,

Itoitz Rodrigo-Jusue, Loughborough University, Loughborough, UK

Pierangelo Rosati, University of Galway, Galway, Ireland

Abstract

The aim of this chapter is to explore the relationship between sport and online hate, via a case study of the work of Tackling Online Hate in Football (TOHIF). In elucidating the challenge of online hate in sport, the chapter showcases how TOHIF is responding to online hate through a range of timely and innovative methodologies and interventions, including accessing large-scale longitudinal empirical datasets, developing tools and techniques to identify and classify online abuse, undertaking over 120 interviews and focus groups with professional male and female footballers from across the UK, a major survey of football fans, interviews with player care professionals, and interviews and educational workshops with sport journalists. The chapter advocates that scholarly research must play a leading role in combatting online hate; initially by deepening the understanding of it, and additionally through shaping the public discourse and education on how to respond to it. We finish this chapter by advocating for some blue sky thinking, and identifying some core areas for future investment.

Introduction

Sport is a valuable social milieu for examining online hate. Over the last two decades, a range of definitions of online hate and online hate speech have emerged. Previously, in Kearns et al. (2023a), we have drawn upon Kilvington's (2021, p. 258) definition, which refers to online hate as 'spreading, inciting, or promoting hatred, violence and discrimination against an individual or group based on their protected characteristics; which include "race", ethnicity, religion, gender, sexual orientation, disability, among other social demarcations'. The European Union's Committee of Ministers' definition (adopted on 20 May 2022) of hate speech provides greater nuance, and includes:

> "all types of expression that incite, promote, spread or justify violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status such as "race", colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation" (Council of Europe, 2022).

Importantly, the EU Council of Ministers note that hate speech varies in terms of severity, some of which may be prohibited or subject to criminal and/or civil or administrative law, and 'other offensive or harmful types of expression which are not sufficiently severe to be legitimately restricted under the European Convention on Human Rights' (Council of Europe, 2022). Thus, hate, as understood for the purposes of this paper, is constituted of acts of abuse (verbal, physical or digital) tied to deeper, underlying systemic oppressions (such as racism, sexism, ableism, etc.).

The aim of this chapter is to explore the relationship between sport and online hate, via a case study of the work of Tackling Online Hate in Football (TOHIF)[1]. TOHIF is an Arts and Humanities Research Council and Irish Research Council funded project, representing a collaboration between Loughborough University, Leeds Beckett University, Sheffield Hallam University, Ulster University (Belfast), and Dublin City University. TOHIF is concerned with framing, tracing, and combating online hate in football. In elucidating the challenge of online hate in sport, the chapter showcases how TOHIF is responding to online hate through a range of timely and innovative methodologies and interventions.

Online hate and sport

Our focus in this chapter is sport; more specifically, major international football tournaments. While the issue of hate is one which concerns sport at all levels, major international tournaments are particularly significant as sites where national, ethnic and gender identities are highlighted, debated and re-shaped across print, broadcast, and digital media. It is our contention that the mediatisation and digitalisation of sporting events position them as flashpoints for online hate (Frandsen, 2014; Kearns et al., 2023a). While one of the media's key advantages is to "bring the mass audience closer" to the individuals and action (Frandsen,

[1] See www.tohif.com

2014, p,535), social media has actively enabled audiences and spectators initiative and spread vitriol.

Inevitably, the conviviality so often expressed through shared identities will, at times, mutate into discriminatory (even violent) behaviour, as a consequence of the passionate rivalries sport promotes (Millward, 2008; Davis, 2015; Kilvington & Price, 2019). Moreover, the increasing pervasiveness of online communication has necessitated we place greater scrutiny on the relationship between sport and expressions of online hate. As many contributions to this volume will attest, media are important ideological vehicles and critical agents in the commercialisation of sport, such that, today, sports events and media are intimately connected (Billings & Wenner, 2017; Black, 2015; Kilvington, 2021; Malcolm & Fletcher, 2017; Stoney & Fletcher, 2020). Thus, the two exist in a symbiotic relationship, reinforcing each other's prominence and popularity (McGillivray & McLaughlin, 2019). A media issue, such as online hate, takes on particular significance and cultural resonance when viewed through the enclave of sport, a totem for identities and a key battleground for cultural disputes. Crucially, instances of online hate in sport are becoming more prominent and prolific.

Online hate and football

Football is the most popular sport in the world, arguably the most popular pastime in the world. It is discussed and promoted across every form of social media. The rise of online hate speech in football is a growing concern, with fans, players, officials, and media figures subject to racist, sexist and homophobic abuse (in addition to many other prejudices). While hate speech and discrimination have always been problems in football, and many other sports, the growth of social media, coupled with football's global popularity, has exacerbated the problem, with spectator hate having largely migrated from the football stands to digital spaces (Kilvington et al., 2023). Due to the pace and scale of development, policy makers, football governing bodies, and anti-hate/discrimination organisations have largely been left playing catch-up (Kilvington & Vester, 2024).

As the most-watched football league in the world, the English Premier League (EPL) has been the subject of much attention from media, policymakers and academic researchers. Kick It Out (2015), England's highest profile anti-discriminatory charity in sport, were the first to commission research to understand the frequency of online abuse in football. During the 2014-15 EPL season, they revealed that there were approximately 134,400 (16,800 per month) discriminatory posts directed at EPL players (39,000) and Clubs (95,000). Further industry commissioned studies, alongside academic research investigating football related social media abuse, has since followed (see Farrington et al., 2015; Kilvington & Price, 2017; Kilvington et al., 2023).

In 2020, the Professional Footballer's Association (PFA), supported by Kick It Out (PFA & Kick it Out, 2020), commissioned data science company Signify to examine targeted, abusive messages sent to 44 high profile current and former male players from England's top four professional divisions. The study covered six weeks of 'Project Restart', the delayed

resumption and conclusion of the COVID-19-impacted 2019-20 season. The study analysed 825,515 incoming messages and found that 43% of the sample had received a targeted, abusive message (i.e., a post aimed directly at them (tagging/hashtags), as opposed to a generic post mentioning a player); 50% of all abusive messages were sent to just three players; and 29% of racist abuse came in the form of emojis. The latter finding was crucial in identifying how emojis represented a blind spot for social media platforms; illustrated by how this content was less likely to be removed (Mubarak et al., 2023).

In collaboration with the Alan Turing Institute, Vidgen et al. (2022) similarly explored the EPL and released a report which tracked abuse on Twitter/X towards players in the EPL across the 2021-22 season. Their machine learning tool found that out of 2.3 million analysed tweets, 59,871 abusive tweets were directed at players, with 68% of players receiving abuse at least once. Similar to the above study, they also found that over half (55%) of posts were positive in nature, with 3.5% being classified abusive. The report found that 50% of all abusive tweets were sent to just 12 players. While these studies and datasets are illuminating, if we are to truly understand the nature and extent of the problem facing football globally, we need to look far beyond elite English men's football. As we highlight below, we are now starting to see more research focusing on online hate in women's football (FIFA & FIFPRO, 2023a), global tournaments (FIFA & FIFPRO, 2023b), and other European leagues (Montesinas-Cánovas et al., 2023; Miranda et al., 2023).

Research focusing on the men's FIFA World Cup 2022 in Qatar (FIFA and FIFPRO, 2023b), sheds some light on online hate in football from a global perspective. The research, which explored Twitter/X and Instagram, identified various triggers for abuse. These included knockout games resulting in increased online hate and threats; players being targeted due to political allegiances and personal circumstances; and in-game flash-point incidents, such as penalty misses or red cards, catalysing hate. The prevalence of abuse and hate was divided against a range of criteria, including: general (26.24%), sexual (17.09%), sexist (13.47%), homophobic (12.16%), and racist (10.70%). Overall, of the 434,000 posts that were flagged by AI and reviewed by humans, 19,636 posts were regarded as abusive, discriminatory, and/or threatening.

FIFA and FIFPRO's (2023a) most recent study into the women's FIFA World Cup 2023 in Australia and New Zealand was the first to comprehensively examine online hate across a global women's tournament. The research examined Twitter/X, Instagram, Facebook, TikTok, and YouTube activity. Following human review of the 102,511 posts that were flagged, 7,085 posts were deemed discriminatory, abusive or threatening, and reported to the respective platform. Twitter/X saw the highest volume of abuse. One in five players at the tournament received targeted online abuse.

It is against this backdrop that the TOHIF project exists. As an interdisciplinary team, consisting of sociologists, marketers, communications scholars, and computer scientists, we are leading a series of workstreams that ensures TOHIF collects and has access to, the largest and most diverse dataset pertaining to online hate and football currently available. The

remainder of this chapter is devoted to these various workstreams; explaining how each is responding to the diverse challenges already identified.

Our approach

1. Scoping the field

To take stock of the academic field of online hate in sport and, therefore football, we undertook a scoping review. Using the database Scopus, we searched all literature published on online hate and sport between 2000-2022. Our initial search identified 406 results, which were eventually whittled down to 41 (see Kearns et al., 2023a). This scoping review was updated, as part of Dublin City University's collaboration with the Fédération Internationale de l'Automobile (FIA), to account for work published between 2022 and end of 2023 (Sinclair et al., 2024).

When conducting the review, we were initially concerned with discovering which sports were most focused on by researchers and how widely this research had been disseminated – thus pointing towards gaps and opportunities. The research field has grown since the mid-2000s: from an initial six journal papers published between 2005 and 2014, a further 31 between 2015 and 2021, to 29 published in one year between 2022 and 2023. Clearly, this rise is linked to the growth of social media itself. In the context of sports-related research, high profile protests such as taking the knee against systemic racism (Duvall, 2020; Black et al., 2023), and campaigns against everyday sexism in sport (e.g., #morethanmean; Antunovic, 2019), have positioned sport at the forefront of wider 'culture wars' (Kearns et al., 2023a; Sinclair et al., 2024).

Early publications (2005 to 2011) focus primarily on message boards and fan forums, before attention in the 2010s shifted to the big social media platforms (Twitter/X, Facebook). Indeed, Twitter/X has been the most popular data source, with 28 of the 66 publications (42.4%) centralising the platform. This can be explained by several factors. Twitter/X operates at a more real-time cadence than other social media platforms and so is particularly suited for live discourses surrounding sporting events. In addition to being a popular medium to discuss sports, it is also used for sharing social and political views. This is enabled by its functionality that enables users to follow those with similar or opposing views using its search engine and hashtags (Kavanagh et al., 2022; Kilvington et al., 2023). Furthermore, up until recent ownership changes, researchers had relatively low-cost access to data through the Twitter API (application programming interface) (Sinclair et al., 2024).

In terms of the location of research, until 2015, publications concentrated exclusively on sport played in the UK and the USA, and has subsequently taken on a more international focus to include Australia, Canada, Russia, Poland, Mexico, Iran, Italy and Brazil. Still, there are only a handful of studies at most in each of these national contexts. Perhaps unsurprisingly, given its status as the most popular sport worldwide, association football, or soccer, was the focus of 24 (38.2%) of the studies, followed by American Football with eight (12.1%) and studies that focused on sport more generally, sharing eight as well.

Racism was the most prominent expression of hate, serving as the focus in 25 of the 66 papers (38%). Misogyny featured in 15 of the publications, the vast majority published from 2020 onwards. Several papers examined an intersection of various types of hate. Of note was the relatively sparse attention paid to other forms of hate, notably disablism and anti-Semitism. The most frequent recipient of hate were athletes themselves (Doidge, 2015; Litchfield et al., 2018; Kavanagh et al., 2019; Duvall, 2020). Often, the abuse of athletes was triggered by a flashpoint event. Other studies illustrated that hate was aimed at an abstract target (i.e., a marginalised group, rather than specific individuals) and highlighted the effect of this on the maintenance of in- and out-group boundaries (Kearns et al., 2023a).

Several recommendations were proposed by both the initial review and the updated version. Notable is the call for much more extensive primary data collection, and particular emphasis was also placed on broadening the focus of sporting and hate 'contexts', developing a research infrastructure that is informed by, and engages with, a much greater number of stakeholders, and that reflects a wider variety of methodologies. The remainder of this chapter identifies how TOHIF has responded to this call.

2. Big data and machine learning

To explore the evolving relationship between sport and online hate, the TOHIF consortium sought to address a few technical challenges, the most significant of which were (a) accessing large-scale longitudinal empirical datasets, and (b) developing tools and techniques to identify and classify online abuse. TOHIF chose the European Football Championships (the Euros) discourse on Twitter (now known as X) as the empirical context. Twitter as chosen for many reasons. First, over the period it was widely used in football discourse across all countries in the Euros by a wide range of actors including teams, athletes, fans, media, and associated journalists, amongst others. Second, Twitter is largely an open network and facilitates the connection, sharing and consumption of content between both acquaintances and strangers (Lynn et al. 2021). Third, over the focal period, a variety of features and functionality were introduced to Twitter that could be used to inform online abuse research. For example, while hashtags (#) emerged in 2007 they were not widely used until after 2009. Hashtags enable Twitter users to identify others with similar and opposing views and form ad hoc and calculated publics around a specific hashtag (Bruns & Burgess, 2011). Similarly, retweets were only introduced in 2009 and image sharing in 2011. Fourth, Twitter's enterprise API platform, GNIP, provided historic access to data with substantial metadata to inform research. This last factor was particularly important as it allowed the project to create substantial multi-level longitudinal datasets that were longitudinal within each tournament and across the 15-year period. Using a set of tournament-related hashtags, over 48 million tweets were ultimately collected for the Euros in all languages using the Twitter Enterprise API, approximately 22 million of which were in English. For each tweet, up to 58 additional data points (e.g. user profiles, date stamps, reply metrics, retweet metrics etc.) were collected.

Table 1 Number of tweets per TOHIF dataset by UEFA European Football Championship

Men's	**Total (all languages)**	**Total (English)**	**Post-processed**
2008*	19,219	11,205	10,903
2012	4,370,308	1,930,192	1,631,242
2016	4,259,499	1,894,293	1,626,095
2020 (2021)	19,851,633	11,021,908	9,506,050
Sub-total	*28,500,659*	*14,857,598*	*12,774,290*
Women's	**Total (all languages)**	**Total (English)**	**Post-processed**
2009*	108,562	39,718	18,271
2013	2,317,550	814,659	149,984
2017	2,411,737	767,475	467,672
2022	15,598,085	5,342,982	2,193,643
Sub-total	*20,435,934*	*6,964,834*	*2,829,570*
Total	**48,936,593**	**21,822,432**	**15,603,860**

*We hypothesise that 2008 and 2009 have lower tweet volumes as hashtags were not widely used at this point.

With a dataset of millions of tweets and billions of data points, advanced tools and techniques are required to identify and classify online abuse. While there has been significant work done in the last decade on developing tools and techniques for classifying hate speech, particularly in recent years with the introduction of large language models (LLMs) based on Transformer architectures, these models are typically trained and fine-tuned on more general language contexts. Football is a unique context possessing its own linguistic idiosyncrasies and terminology with cultural and regional variations (Billings, 2016; Lavric et al. 2008). As such, merely applying existing tools - even those fine-tuned for hate speech - such as BERT-Hate (Kralj Novak et al. 2022) and RoBERTa Offensive (Barbieri et al. 2020), can result in significant false positives. Unfortunately, there were few suitable dictionaries and annotated datasets of online abuse in international football tournaments for training and testing machine learning and deep learning models.

Therefore, a second workstream of TOHIF, supplemented by the FIA United Against Online Abuse (UAOA) project, involves developing resources and models for identifying specific tools and techniques to support the automated classification of online abuse and hate speech. This involves five discrete activities. First, we have and continue to develop discrete vocabularies of terms, word stems, and phrases for detecting potential hate speech and classes of hate speech (e.g., racism, homophobia, sexism, xenophobia, islamophobia, ableism, and so on). These general vocabularies are then used to generate context- and class-specific datasets from our tournament data which are subsequently annotated using pre-trained LLMs and human coders. Following further pre-processing, these annotated or labelled datasets are used to fine-tune and evaluate machine learning and deep learning models for identifying and classifying specific classes of online abuse and hate speech in the Euros. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. To further optimise the models, we analyse the impact of different input text on model classification using explainable artificial intelligence (XAI). Once acceptable performance levels have been achieved, typically over 95%, we undertake further pre-processing on the master Euros dataset

and operationalise a hybrid classification system that combines class-specific keyword matching and LLM components to identify and classify the entire Euros dataset for the focal hate speech class. Identifying and classifying online abuse and hate speech should not be seen as an end point to the analytics task. Once such tweets have been identified, as per Lynn & Rosati (2022) we use exploratory data analysis and a combination of descriptive, user, content, and network analytics to identifying events or phenomena of interest or address specific research questions raised by other parts of TOHIF.

We believe the technical work under TOHIF, which is being supplemented and continued under the FIA UAOA project, directly benefits several stakeholders. Our intention is to make the custom vocabularies, annotated datasets, preprocessing recommendations, and models for each class available as open data, to the extent permitted under existing licenses, and through publications in due course. These resources, tools, and techniques advance the state of the art in online abuse and hate speech detection, and provide valuable datasets and methodological tools for researchers in computer science, sports, and social sciences. The EU Digital Services Act (DSA) introduces regulations to create a safer online environment by imposing specific responsibilities and accountability measures for digital services, including social media platforms (Cauffman & Goanta, 2021). It requires online platforms to effectively address and remove illegal content, including hate speech, and to be transparent about their content moderation practices (Cauffman & Goanta, 2021). By advancing research on different classes of hate speech, online platforms, such as Facebook, TikTok and X, and specialist hate speech detection technology developers, such as Arwen and Signify, can improve the performance of their content moderation services and models. Additionally, policymakers, sports organisations, and advocacy groups can use the insights from identifying and understanding different linguistic patterns, events, users, and networks in different hate speech classes to inform policy and targeted interventions to combat online abuse and hate speech on social media. This includes more effective counter-messaging, awareness campaigns, educational programmes, and support services for victims of online abuse.

3. Players and player care professionals.

On the one hand, social media for elite sportspeople opens a world of opportunities leading to brand deals, increased income and adoration from fans (Kavanagh et al., 2019, 2022, 2023). On the other hand, social media can destroy careers or cause dangerous levels of stress. It is frequently said that online abuse is simply the price you have to pay when you are a professional athlete. Such trivialisation and, to a certain extent, dehumanisation, must be challenged. The lack of empathy towards athletes who experience online hate is perhaps the result of a lack of research and insight into athlete responses to hate. As reinforced in Kearns et al. (2023a) and Sinclair et al. (2024), there is a significant lack of primary, qualitative research with professional athletes. As a result, we actually know very little about the nature and extent of the hate they receive, how they respond to it, and what its impact is on their identity, mental health, and wellbeing.

TOHIF is responding to this through working with professional footballers. At the time of writing, 163 players, including current and former male and female players, representing Clubs

from the Premier League, Championship, League One, League Two, Women's Super League and Women's Championship had participated in one-to-one semi-structured interviews and/or focus groups. Given the relative inaccessibility of professional athletes – footballers particularly – this sample is unprecedented. Our conversations with players have reinforced that, despite in-game trigger events often leading to an increase in online hate, there are other factors too which should be explored. For example, players who challenge discrimination and use their voice and platforms to influence positive change are more likely to receive online hate and abuse in return (Kilvington, 2021; PFA, 2020; FIFA & FIFPRO, 2023b). Moreover, players are also facing various other 'harms', ranging from online harassment, bribery, death threats (aimed at both themselves and their loved ones), and instances of mis/disinformation and fake news which may have professional and personal implications.

It is, therefore, crucial that players are supported and protected by the clubs (their employer), leagues and governing bodies, and provided with comprehensive structures and guidance on how to anticipate and cope with abuse. While professional football clubs, especially in the top leagues, are likely to have safeguarding and player welfare officers on hand to support players (see Hickey & Roderick, 2023), many clubs are under-resourced, and personnel lack specialist knowledge around online hate, its impacts, and how to respond (Kilvington and Price, 2017). To date we have interviewed six player care educators and professionals from across the EPL, English Football League, and Women's Super League, with the aim of better understanding the level of education provision available, as well as existing structures for responding. Through collaboration with these player care professionals, TOHIF is developing a range of open access educational resources around online hate, aimed at players and club staff, in order to transform this insight into action.

4. Fans

Both Kearns et al. (2023a) and Sinclair et al. (2024) identified that the bulk of research focuses on online hate perpetuated by 'fans'. However, currently, little is known of the characteristics (age, gender, ethnicity, political affiliations, socio-economic status, family status etc.,) or motivations of fans who engage in hate speech. This is mainly because a) a dearth of primary research into online hate has been undertaken with fans; b) studies that have utilised secondary data have either scraped or screenshot posts directly from social media, without accounting for who has posted the comment; and c) it is often difficult to verify who owns and posts from social media accounts. Nevertheless, without these insights, there is a risk of viewing fans in a narrow binary of active abusers and passive observers. Little research has been conducted into online hate targeted at fans or into fans' role in tackling the problem.

To respond to this TOHIF is currently surveying football fans on their experiences of online hate as perpetrators, targets and activists, with the aim of better understanding the contexts underpinning hate, the motivations for spreading hate, and the impact of occupying spaces where hate openly proliferates. This survey will help ensure the voices of fans are heard when it comes to devising actions to address the problem of online hate in football.

5. Journalists and media organisations

As Kearns et al. (2023a) have previously identified, the bulk of research into online hate in sport focuses on incidents of players receiving hate from fans. However, research into online hate in other contexts has begun to explore the problem of online hate directed at journalists, albeit more slowly than any expansion in the scale of the problem (e.g., Binns, 2017; Stahel and Schoen, 2019; Obermaier, 2023;). The experiences of sports journalists, however, have been largely neglected in this space (see Kilvington & Price, 2022).

TOHIF has begun to fill this gap: to date, we have interviewed 21 Irish and UK sports journalists; asking them to reflect on their experiences of online hate and abuse. Findings reveal that: first, sports journalism is a distinctive and challenging profession because significant emotional work is undertaken. This is largely due to the need to forge bonds and connections with the readership, and thus, is accentuated when online abuse occurs. Journalists are deeply affected by this abuse, personally and professionally; second, online abuse is now so ubiquitous as to be habitually accepted, and it has also obscured the distinction between public and private spaces; third, in the absence of formal institutional supports around online abuse, or even an informal code of practice about how to manage this, sports journalists have been compelled to develop their own emotional strategies, including self-censorship. They see this as unwelcome, but a necessary protective mechanism. The findings raise a series of important practical and existential questions about the sustainability of the profession if journalists continue to rely on their own emotional devices to manage online abuse. In addition to these interviews, TOHIF has conducted three action-based workshops with UK-based sports journalists at a leading global news organisation. Findings from these workshops will be used to help shape this organisations' response to online hate going forwards, education for its employees, and support systems for recipients of hate and abuse.

6. Coordinated hate campaigns

Co-ordinated hate campaigns remain underexplored. The majority of publications to date address cases of what might be termed 'spontaneous hate', arising from the interaction between a flashpoint incident and societal and cultural prejudices, whether latent or overt. Some research has explored the coordinated actions of some members of a supporter base in perceived defence of their team's reputation and/or identity (e.g. McLean et al., 2017). However, only Glathe and Varga (2018) examine the role of organised extremist ideologies in shaping incidents of online hate in sport. Given the recent mainstream media attention into the leveraging of sport by far-right groups (Aroas, 2019), much more could be gleaned through detailed analysis of how such groups enact and exacerbate incidents of online hate.

TOHIF is actively contributing to this agenda through the work of Black et al. (2023) and Doidge et al. (*forthcoming*), which centres online discourses surrounding players taking the knee during the men's 2020 UEFA European Championship. Black et al's analysis helps us to understand how discourses of hate and far-right narratives were created, spread and gained

traction on social media, identifying how alt-right and far-right conspiracy theories were constructed and circulated on Twitter/X in the context of football. The article argues that sport, and particularly football, has become a popular vehicle for spreading conspiracy theories. Accordingly, alt-right conspiracies have frequently denounced Black Lives Matter (BLM) and "take the knee" protests by suggesting that the movement and its symbols represent little more than a guise for promoting anti-White sentiment.

Black et al., revealed how much of the vitriol directed at BLM was fuelled by assertions that the movement's fundamental goal was to advance a Marxist political agenda. By placing sport in the broader context of political discussions concerning the take the knee protest, we can begin to emphasise the specific mechanisms through which the alt-right generates increased prominence in public discourse. Through an analysis of 1,388 tweets during the tournament, Black et al., demonstrated that the alt-right used the take the knee protest to attack BLM and the wider 'woke agenda'. The online rebuttal to the take the knee protest endorsed an anti-White racist position, while also highlighting contemporary feelings of White victimhood and resentment. This sheds further light on the nexus between far-right ideologies and football within the context of major sporting events, underscoring the pivotal role of conspiracy in the broader dissemination of alt-right ideology across popular culture.

Drawing on theories of aversive racism and colour-blindness, which stress the invisibility of contemporary racism, Doidge et al. (*forthcoming*) showed how subtle racism (as contrary to overt and/or blatant expressions of racism) and colour-blindness (apparently non-racial 'new racism' practices, see Bonilla-Silva, 2014) were reinforced in online discussions around 'taking the knee'. While public figures and the media only focused on condemning highly visible racist abuse directed at Black English players after losing the final to Italy, based on an analysis of 6,850 English language tweets published on Twitter/X over the duration of the tournament, Doidge et al. (*forthcoming*) revealed how more subtle forms of racism were reinforced in online discussions around 'taking the knee'. Moreover, they also emphasised how Twitter/X users will develop sophisticated counter-narratives and strategies to challenge arguments against anti-racist activism in football. The research revealed how some factors (i.e., key moments in the tournament, political elites' discourses, and trends of success and failure) shaped online discussions on, and expressions of, (anti)racism. Doidge et al. (*forthcoming*) emphasised the necessity of investigating the complex temporality of racism in public conversations, which is most often ignored in research on online hate.

Through Kilvington et al. (2023), TOHIF has also examined existing organised campaigns from within the football community to challenge online hate. These included analysing the efforts of anti-discrimination group Kick it Out, specifically its Klick it Out campaign, and the ineffectiveness of social media boycotts, such as those facilitated in 2019 and 2021, and police deterrents, epitomised by the launch of the United Kingdom's Football Policing Unit (UKFPU). The overarching argument is that awareness raising – through high profile resistance campaigns is a necessary first step. However, the conversations, collaborations, and actions that follow, make the real impact.

Conclusion and future directions

All the evidence suggests that online hate will continue to grow, mutate, and diversify. The relative impotence of most social media platforms, and indeed sports' governing bodies, to act on the root causes of hate is a case in point. We strongly believe that scholarly research must play a leading role in combatting online hate; initially by deepening the understanding of its causes, dynamics, and consequences, thereby informing more effective policies and interventions; and additionally through shaping the public discourse and education on how to respond to it. Thus, the academic community is uniquely placed to make a significant impact on this space. We finish this chapter by advocating for some blue sky thinking, and identifying some core areas for future investment.

Collaboration

TOHIF is the first study to examine online hate longitudinally across a 14-year period, and across multiple sporting events. From a methodological perspective, the evolving and subjective nature of language, varying definitions of hate speech, and the complexity of online interactions, particularly on social media, makes this kind of large-scale quantitative research challenging. Even where access to data is available, researchers may face scale problems (Lynn & Rosati, 2022). While the number of posts featuring online hate speech may be relatively small, identifying hate speech from very large datasets, including multiple languages (each with their own cultural nuances and colloquialisms), such as those associated with mega sporting events like a football World Cup, may require advanced techniques and specialist skills rarely evident among those who have traditionally examined online hate in sport – that is, sociologists of sport and communication scholars. Thus, to develop more nuanced research, there is a great need for much more large-scale, international, interdisciplinary collaboration, including scholars from psychology, sociology, sports studies, linguistics, and data science, for nuanced research.

Diversify platforms

As we have emphasised throughout, much of the academic work in this space has focused on only a small number of social media platforms, and with a significant reliance on Twitter/X. It is important to note that, as in the past, the online platform context is likely to change moving forward (Sinclair et al., 2024). Firstly, with the change in ownership of Twitter, research using the Twitter API is significantly more expensive and, as such, may be prohibitive for research at scale. Additionally, changes in moderation policies at X have resulted in a reported increase in online hate speech in the platform (Frenkel & Conger, 2022). Secondly, new online platforms, such as TikTok, are increasingly the primary information source and social network for younger consumers. TikTok has only very recently released access to its research API, and it is not yet known how accessible data will be. We should not ignore the continued relevance of blogs and fan forums (Kearns et al., 2023b). At the same time, a number of niche platforms have emerged for users with more extreme views - e.g., Trust Social, Gab etc., (Jasser et al.,

2023). These niche platforms place less emphasis on content moderation compared with the 'main' platforms. These platforms can act as breeding grounds for hateful discourses and are often seen as 'lifeboats' for users who have been blocked on Instagram and Facebook (Kearns et al., 2023a). Similarly, there is a need for research on encrypted messaging apps, such as WhatsApp, Telegram and Signal. Platforms with specific geographical coverage (e.g., WeChat) might also be considered. Fourthly, no research on the intersection of the dark net, hate speech and sports was identified in the literature. Finally, new analytical techniques are enabling analysis of multimedia, including audio, images and video. This opens new research opportunities on content from platforms, such as YouTube and Instagram, amongst others. These new techniques will open potentially massive datasets of content that has previously been ignored due to lack of automation (Sinclair et al., 2024).

Widening the coverage of online hate speech

Much of the research has focused on online hate in the context of a relatively small number of protected characteristics – namely 'race', ethnicity, gender, and sexual orientation. The narrow focus is problematic for developing universal measures to combat online hate across all identity markers. This necessitates a widening of the coverage to include under-researched types of online hate speech, specifically transphobia, anti-Semitism, Islamophobia, xenophobia, sectarianism, and ableism, and the relationship between different types of online hate speech. The growing discourse concerning the participation of transgender athletes in professional sport will likely lead to a growth of research in that space. We would also expect to see more research concerning Islamophobia considering the men's FIFA World Cup 2022 took place in Qatar, and the high-profile investment of Middle Eastern countries, such as Saudi Arabia, at the top level of sport will likely influence public perceptions of the region further (see Crossley & Woolf, 2024).

Action(ing) research

Much of the academic work into online hate and sport is at the understanding phase. There is a need to accelerate the transition from understanding hate to challenging and combatting it. This movement necessitates that we ask questions about what we are doing with our data and insight. Working in tandem with policymakers, online platforms, sports federations, and sports organisations, we need to test, assess, and refine different measures and interventions to combat online hate speech. Only then will academic work have the impact it so desires.

References

Antunovic, D. (2019). 'We wouldn't say it to their faces': Online harassment, women sports journalists, and feminism. *Feminist Media Studies*, *19*(3), 428-442.

Aroas, C. (2019). *Proud boys and skinheads: MLS faces an incursion from the far-right*. Available at: https://www.theguardian.com/football/2019/mar/06/proud-boys-andskinheads-mls-faces-an-incursion-from-the-far-right

Barbieri, F., Camacho-Collados, J., Neves, L. & Espinosa-Anke, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. arXiv preprint arXiv:2010.12421.

Billings, A.C. (Ed.) (2016). *Defining sport communication*. Routledge.

Billings, A.C., & Wenner, L. (Eds.) (2017). *Sport, media and mega-events.* Routledge.

Binns, A. (2017). Fair game? Journalists' experiences of online abuse. *Journal of Applied Journalism & Media Studies*, *6*(2), 183–206.

Black, J. (2015). Portraying Britain's Past: English National Newspaper Coverage of the 2012 London Olympic Ceremonies. In G. Poynter, V. Viehoff, & Y. Li, (Eds.), *The London Olympics and Urban Development* (pp. 320-334). Routledge.

Black, J., Fletcher, T., Doidge, M., Kearns, C., Kilvington, D., Liston, K., Lynn, T., Rosati, P., & Sinclair, G. (2023). 'Let the tournament for the woke begin!': Euro 2020 and the reproduction of Cultural Marxist conspiracies in online criticisms of the 'take the knee' protest. *Ethnic and Racial Studies*. DOI: 10.1080/01419870.2023.2263069

Bruns, A. & Burgess, J. (2011). The use of Twitter hashtags in the formation of ad hoc publics. In *Proceedings of the 6th European consortium for political research (ECPR) general conference 2011* (pp. 1-9). The European Consortium for Political Research (ECPR).

Cable, J., Kilvington, D., & Mottershead, G. (2022). 'Racist behaviour is interfering with the game': Exploring football fans' online responses to accusations of racism in football. *Soccer & Society*, *23*(8), 880-893.

Cauffman, C., & Goanta, C. (2021). A new order: The digital services act and consumer protection. *European Journal of Risk Regulation*, *12*(4), 758-774.

Council of Europe. (2022). *CM/Rec(2022)16*. Available at: https://rm.coe.int/0900001680a67955

Crossley, S., & Woolf, A. (2024). 'Fog on the Tyne?': The 'common-sense' focus on 'sportswashing' and the 2021 takeover of Newcastle United. *International Journal of Sport Policy and Politics*, *16*(2), 307-322.

Davis, L. (2015). Football fandom and authenticity: A critical discussion of historical and contemporary perspectives. *Soccer & Society*, *16*(2–3), 422–436.

Doidge, M. (2015). 'If you jump up and down, Balotelli dies': Racism and player abuse in Italian football. *International Review for the Sociology of Sport*, *50*(3), 249–264.

Duvall, S.S. (2020). Too famous to protest: Far-right online community bonding over collective desecration of Colin Kaepernick, fame, and celebrity activism. *Journal of Communication Inquiry*, *44*(3), 256-278.

Farrington, N., Hall, L., Kilvington, D., Price, J., & Saeed, A. (2015). *Sport, racism and social media*. Routledge

FIFA., & FIFPRO. (2023a). *Social media protection service – FIFA Women's World Cup Australia & New Zealand 2023: Full tournament analysis', FIFA Threat Matrix*. Available at: https://digitalhub.fifa.com/m/36bb7afdd56112ab/original/FIFA-Social-Media-Protection-Service-FIFA-Women-s-World-Cup-2023-tournament-analysis.pdf

FIFA., & FIFPRO. (2023b). *Social media protection service – FIFA World Cup Qatar 2022: Tournament analysis, FIFA Threat Matrix.* Available at: https://fifpro.org/media/zp2fcida/fifa-smps-threat-matrix-release-world-cup-tournament-report-v7.pdf

Frandsen, K. (2014). Mediatization of sports. In K. Lundby (Ed.), *Mediatization of communication* (pp.525-543). Mouton de Gruyter.

Frenkel, S., & Conger, K. (2022). *Hate speech's rise on Twitter is unprecedented, researchers find*. Available at: https://www.nytimes.com/2022/12/02/technology/twitter-hate-speech.html

Glathe, J., & Varga, M. (2018). Far-right fan culture in Russia: The politicisation of football hooligans on Russian social media. *Journal Fur Entwicklungspolitik*, *34*(2), 20–49.

Hickey, C., & Roderick, M. (2023). Career transitions from the English Premier League: Cooling out the mark with possible selves. *International Review for the Sociology of Sport*, *58*(6), 1050-1072.

Holton, A., Bélair-Gagnon, V., Bossio, D., & Molyneux, L. (2021). 'Not their fault, but their problem: Organizational responses to the online harassment of journalists. *Journalism Practice*, *17*(4), 859-874.

Jasser, G., McSwiney, J., Pertwee, E., & Zannettou, S. (2023). 'Welcome to# GabFam': Far-right virtual community on Gab. *New Media & Society*, *25*(7), 1728-1745.

Kavanagh, E. J., Litchfield, C., & Osborne, J. (2023). Critiquing the social media scholarship in sport studies: Social media and athlete welfare. *International Journal of Sport Communication*, pp.1-24.

Kavanagh, E., Litchfield, C., & Osborne, J. (2022). Social media, digital technology and athlete abuse. In J. Sanderson (Ed.) *Sport, social media, and digital technology* (pp.185-204). Emerald.

Kavanagh, E., Litchfield, C., & Osborne, J. (2019). Sporting women and social media: Sexualization, misogyny, and gender-based violence in online spaces. *International Journal of Sport Communication*, *12*(4), 552-572.

Kavanagh, E., Jones, I., & Sheppard-Marks, L. (2016). Towards typologies of virtual maltreatment: Sport, digital cultures and dark leisure. *Leisure Studies, 35*(6), 783–796.

Kearns. C., Sinclair. G., Black. J., Doidge. M., Fletcher. T., Kilvington. D., Liston. K., Lyn. T., & Rosati. P. (2023a). Scoping Review into Academic Research Conducted into Online Hate and Sport. *Communication & Sport, 11*(2), 402-430.

Kearns, C., Sinclair, G., Black, J., Doidge, M., Fletcher, T., Kilvington, D., Liston, K., Lynn, T., & Santos, G. L. (2023b). 'Best run club in the world': Manchester City fans and the legitimation of sportswashing? *International Review for the Sociology of Sport*. https://doi.org/10.1177/10126902231210784

Kick It Out. (2023). *Reporting statistics*. Available at: https://www.kickitout.org/reporting-statistics

Kick It Out. (2015). *Kick It Out unveils findings of research into football-related hate crime on social media*. Available at: http://www.kickitout.org/news/kick-it-out-unveils-findings-of-research-into-football-related-hate-crime-on-social-media/#.VaT58RZChz8

Kilvington, D., & Vester, T. (2024). *Tackling football related hate speech and far-right narratives online.* Unpublished report prepared for Football Against Racism in Europe. FARE.

Kilvington. D., Black. J., Doidge. M., Fletcher. T., Kearns, C., Liston. K., Lynn. T., Sinclair, G., & Rosati. P. (2023). A critical analysis of past and present campaigns to challenge online racism in English professional football. In I. Awan., & I. Zempi (Eds.), *Hate crime in football* (pp.83-102). Bristol University Press.

Kilvington, D., & Price, J. (2021). The beautiful game in a world of hate: Sports journalism, football and social media abuse. In R. Domeneghetti (Ed.), *Insights on reporting sports in the digital age: Ethical and practical considerations in a changing media landscape* (pp.104-119). Routledge,

Kilvington, D. (2021). The virtual stages of hate: Using Goffman's work to conceptualise the motivations for online hate. *Media, Culture and Society*, *43*(2), 256–272.

Kilvington, D., & Price, J. (2019). Tackling social media abuse? Critically assessing English football's response to online racism. *Communication & Sport*, *7*(1), 64–79.

Kralj Novak, P., Scantamburlo, T., Pelicon, A., Cinelli, M., Mozetič, I., & Zollo, F. (2022). Handling disagreement in hate speech modelling. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 681-695). Springer International Publishing.

Lavric, E., Pisek, G., Skinner, A., Stadler, W. (Eds.) (2008). *The linguistics of football*. Narr Francke Attempto Verlag.

Litchfield, C., Kavanagh, E., Osborne, J., & Jones, I. (2018). Social media and the politics of gender, race and identity: The case of Serena Williams. *European Journal for Sport and Society*, *15*(2), 154–170.

Lynn, T., Rosati, P. & Murphy, B. (2021). Does size matter? Non-market social and political engagement by accounting firms in the #Brexit discourse on Twitter. Accounting Forum *45*(1), 58-84.

Lynn, T., & Rosati, P. (2022). Social media research using big data–types, techniques, and technologies. In A. Hanlon., & T.L. Tuten (Eds.), *The SAGE handbook of social media marketing* (pp.91-109). Sage.

Malecki, W.P., Kowal, M., Dobrowolska, M., & Sorokowski, P. (2021). Defining online hating and online haters. *Frontiers in Psychology*, *12*, 744614.

Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, *20*(6), 930-946.

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, *22*(2), 205-224.

McGillivray, D., & McLaughlin, E. (2019). Transnational digital fandom: Club media, place, and (networked) space. In S. Lawrence., & G. Crawford (Eds.), Digital football cultures: Fandom, identities and resistance (pp.30-46). Routledge.

McLean, S., Wilson, A., & Lee, E. (2017). The whiteness of redmen: Indigenous mascots, social media and an antiracist intervention. *Australasian Journal of Information Systems*, *21*. https://doi.org/10.3127/ajis.v21i0.1590

Millward, P. (2008). Rivalries and racisms: 'closed' and 'open' Islamophobic dispositions amongst football supporters. *Sociological Research Online*, *13*(6), 1–17.

Miranda, S., Gouveia, C., Di Fátima, B., & Antunes, A.C. (2023). Hate speech on social media: Behaviour of Portuguese football fans on Facebook. *Soccer and Society*, *25*(1), 76–91.

Montesinos-Canovas, E., Garcıa-Sanchez, F., Garcıa-Dıaz, J.A., Alcaraz-Marmol, E., & Valencia-Garcıa, R. (2023). *Spanish hate-speech detection in football*. Available at: https://rua.ua.es/dspace/bitstream/10045/137094/1/PLN_71_01.pdf

Mubarak, H., Hassan, S., & Chowdhury, S. A. (2023). Emojis as anchors to detect Arabic offensive language and hate speech. *Natural Language Engineering*, *29*(6), 1436-1457.

Obermaier, M. (2023). Occupational hazards: Individual and professional factors of why journalists become victims of online hate speech. *Journalism Studies*, *27*(7), 838-856.

PFA., & Kick it Out. (2020). *AI research study: Online abuse and project restart*. Available at: https://www.thepfa.com/news/2020/10/21/pfa-charity-ai-research-study#:~:text=During%20the%20six%20weeks%20of,during%20the%20study%20was%20racist.

Sinclair, G., Kearns, C., & Lynn, T. (2024) *FIA review of scholarly publications on online hate in sport*. https://unitedagainstonlineabuse.com/publications/fia-review-of-scholarly-publications-on-online-hate-in-sport/

Stahel, L., & Schoen, C. (2019). Female journalists under attack? Explaining gender differences in reactions to audiences' attacks. *New Media & Society*, *22*(10), 1849-1867.

Stoney, E., & Fletcher, T. (2020). 'Are fans in the stands an afterthought?' Sports events, decision-aid technologies and the Television Match Official in rugby union. *Communication & Sport, 9*(6), 1008-1029.

Vidgen, B., Chung, Y.L., Johansson, P., Kirk, H.R., Williams, A., Hale, S.A., Margetts, H., Röttger, P., & Sprejer, L. (2022). *Tracking abuse on Twitter against football players in the 2021–22 Premier League season.* Available at: https://www.turing.ac.uk/sites/default/files/2022-08/tracking_abuse_on_twitter_against_football_players_web.pdf