

Institution: City, University of London (City)		
Unit of Assessment: C21 Sociology		
Title of case study: Informing Public and Policy Debate on Social Media Dis/misinformation: Twitter bots, Echo-chambers, and Hyper-partisan News in the Brexit Debate		
Period when the underpinning research was undertaken: 2016 – ongoing		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s): Dan Mercea Marco Bastos	Role(s) (e.g. job title): Reader in Media and Communication Senior Lecturer in Media and Communication	Period(s) employed by submitting HEI: 1st October 2013 - present 15th August 2016 - present
Period when the claimed impact occurred: 2016-2021		
Is this case study continued from a case study submitted in 2014? No		
<p>1. Summary of the impact (indicative maximum 100 words) Bastos and Mercea demonstrated for the first-time how a network of social media ‘bots’ – automated accounts – were deployed on Twitter in the 2016 Brexit referendum campaign to artificially amplify electoral messages. Their research received extensive media coverage, stimulating public debate on the strategic electoral deployment and removal of bot networks. The coverage drew the attention of the UK Parliament to the research, which influenced parliamentary inquiries and subsequent legislation proposed by the UK government on the responsibilities of social media companies to tackle disinformation. Used by security services internationally to develop responses to state-sponsored dis/misinformation campaigns, the research led to professional engagement with Twitter, which supported follow-on research by the authors.</p>		
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>Recent City research is making transparent the ‘darker side’ or potential harms posed by social media including the use of algorithms and automated non-human accounts (‘bots’) to game the social media attention economy. [3.1,3.4, 3.6] This includes ‘weaponising’ social media platforms for disinformation campaigns (that is, the manipulation of information that purposefully aims to mislead or deceive) and interference in political activities such as elections with wide-ranging consequences for democracies, public policy, and ethical implications. [3.1,3.5,3.6]. Bastos and Mercea’s research was one of the first peer-reviewed independent investigations into the relation of geographic location, in this case of Twitter users, and the circulation of social media content during a national political campaign, namely the 2016 UK referendum on EU membership (Brexit). [3.1-3.6] An unexpected finding from their analysis was the uncovering of a ‘Twitterbot’ network active during the campaign that was then mostly deleted after the ballot closed. Bots are automated non-human accounts designed to carry out specific tasks online. Through a combination of user activity metrics and an examination of temporal posting patterns that went beyond then conventional frequentist approaches to the study of bots, the research showed how bots acted as ‘sock puppets’, that is, false online identities used to voice opinions and manipulate public opinion. Notably, through rapid retweet cascades, the Twitter botnet was able to seed hyper-partisan content, contributing to an ‘echo chamber’ effect [3.1,3.2]. This discovery was the first research-based evidence in the public domain of a Twitterbot network at play during a UK political campaign.</p> <p>The research focused on the two-week period before and after the EU referendum vote, that is between 10 June and 10 July 2016. Around ten million tweets associated with the referendum were collected and over 800,000 unique users identified. The chosen methodological approach allowed the authors to separate out bots and uncover a botnet (a network of internet-connected</p>		

devices) comprising of 13,493 accounts that tweeted on the referendum, only to disappear from Twitter shortly after the referendum polling stations closed (detailed methodological and ethical considerations were addressed in 3.1). It is important to note that 'sock puppet' bots breach the terms of service of networking sites like Facebook and Twitter.

The botnet was effective at rapidly pushing out user-generated 'hyper-partisan' rather than 'fake' news that simplifies and spectacularises political news stories which, in turn, feed into polarised identities. The research evidenced for the first time how the botnet was subdivided into specialised subnetworks dedicated to 'echoing' through retweets content created either by other bots (driven by algorithms) or humans. [3.1-3.6]

Exploring the geographic patterns of Twitter activity related to the Brexit referendum, within their wider project, the authors were able to show how ideologically polarised 'echo-chambers' on Twitter mapped onto geographically-situated social networks. [3.2] They designed a new approach for algorithmically identifying user location and political affiliation and, through a series of randomized tests, revealed a relation between echo-chamber communication and geographic distance. Their analysis found that supporters of the Leave campaign were more likely to communicate in 'echo chambers' that were associated with geographic proximity. The opposite relationship characterised the Remain campaign, on Twitter.

Reflecting on their research [3.4], Bastos and Mercea warn of the increasing 'weaponisation' of social media platforms and use of algorithms in which bots feature to try and game social media attention. However, their research finds no evidence to support the notion that bots had substantially altered the Brexit debate on Twitter. [3.4] Collectively, these research insights shed new light on how dis/misinformation campaigns can leverage social media features and network effects to rapidly scale up the dissemination of polarising content among susceptible publics. [3.5] The research demonstrates the need for ethical and transparent public research into social media dis/misinformation that maintains public standards of accountability [3.5, 3.6], what Bastos and Mercea refer to as making "the invisible visible" [3.4, p.2].

3. References to the research (indicative maximum of six references)

3.1 Bastos, M. and Mercea, D. (2017/2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*. 37 (1), 38-54. <https://doi.org/10.1177%2F0894439317734157> [Note: first published online in 2017]

3.2 Bastos, M., Mercea, D., and Baronchelli, A. (2018). The geographic embedding of online echo chambers. Evidence from the Brexit campaign. *PLoS One*, 13 (11). <https://doi.org/10.1371/journal.pone.0206841>

3.3 Bastos, M. and Mercea, D. (2018). Parametrizing Brexit: Mapping Twitter Political Space to Parliamentary Constituencies. *Information, Communication & Society*, 21 (7) 921-939. <https://doi.org/10.1080/1369118X.2018.1433224>

3.4 Bastos M. and Mercea, D. (2018) The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign. *Phil. Trans. R. Soc. A* 376: 20180003, 1-12. <http://dx.doi.org/10.1098/rsta.2018.0003>

3.5 Mercea, D. and Bastos, M. (2019) Brexit Tweets and the Polarised Terrain of Dis/misinformation. *Science in Parliament* 75 (2), 32-35. <https://www.scienceinparliament.org.uk/publications/science-in-parliament/>

3.6 Walker, S., Mercea, D., Bastos, M. (2019) The Disinformation Landscape and the Lockdown of Social Platforms. *Information, Communication & Society*, 22(11), 1531-1543. <https://doi.org/10.1080/1369118X.2019.1648536>

4. Details of the impact (indicative maximum 750 words)

The discovery of the Twitterbot network (3.1) and the public policy implications arising from this, had significant impact and reach in policy and public discourses about the role of social media. This falls into three areas of impact: (1) contributing to UK Parliamentary inquiries into the role of social media networks in dis/misinformation and the resulting changes in public policy leading to draft legislation and developing regulatory frameworks; (2) enabling national and international security organisations to form an understanding of botnets and the threats posed by social media

to the integrity of democratic institutions; and (3) bringing these issues into the public spotlight and making ‘visible the invisible’ how social media platforms can be abused and social media companies operate. The public attention prompted Twitter to engage professionally with the research and ultimately to support its further development by the authors.

4.1 Informing UK Parliamentary inquiries and policy in relation to social media platforms

Bastos and Mercea shared their Twitterbot research exclusively with BuzzFeed News. [5.1] The resulting article appeared on 20 October 2017 and explained in detail the findings and implications of the City research. Extensive media attention followed, leading to multiple requests for the authors to explicate their research to multiple beneficiaries, i.e. policy makers, security agencies, media, and the public.

The BuzzFeed article included an interview with Damian Collins MP, the then Chair of the House of Commons Digital, Culture, Media and Sport (DCMS) Select Committee which was holding an inquiry into “Disinformation and ‘Fake News’”. The City research on Twitterbot networks (3.1) was submitted to the inquiry and prompted Collins to take the unusual step of writing to Twitter UK (19 October 2017) and the Twitter CEO Jack Dorsey (14 December 2017, 25 January 2018) specifically raising the issue of the 13,493 ‘bots’ identified in City’s research and asking for more information about the extent to which there had been “interference” in the UK’s democratic process. [5.2, pages 85-86] Twitter later responded directly to the City research in a letter to Collins from the Head of Public Policy (Twitter UK) on 19 January 2018 confirming that Twitter had “conducted analysis on the dataset that City University provided in order to gather the specific information required to respond to your individual questions in full”. [5.2, p. 85] The City research was further discussed in the oral evidence given to the Committee on 8 February 2018 (held in Washington DC) by Twitter’s Director of Public Policy and Philanthropy (U.S. and Canada) and Head of Public Policy (Twitter UK) [5.2, p. 101, Questions 479-568]. City is specifically mentioned in Q484 where Twitter acknowledged that academics do not have access to “the full picture” of user data. [5.2, p. 78] The correspondence and oral evidence can be found in the Committee’s Interim report published on 24 July 2018. [5.2]

The Committee’s Final Report on Disinformation and ‘Fake News’ [5.3] was published on 18 February 2019 and its analysis and recommendations directly influenced the Government’s Online Harms White Paper (WP) (published for consultation in April 2019). [5.4, paragraph 1.25] The WP makes explicit the Government’s concern about disinformation and alludes to a key finding from City’s research (3.1, 3.2) when it says “social media platforms use algorithms which can lead to ‘echo chambers’ or ‘filter bubbles’, where a user is presented with only one type of content instead of seeing a range of voices and opinions”. [5.4, paragraph 4] In setting out the criteria for a ‘duty of care’ with respect to disinformation [5.4, paragraph 7.27-28] there is again clear resonance with the underpinning research (3.1-3.5) when it is proposed to counter the ‘echo chamber’ by social media companies making clear to users when they are dealing with automated accounts (i.e. ‘bots’), ensuring that automated dissemination is not abused and promotes diverse news content. [5.4, section 7.28] The Government’s response to the White Paper consultation (December 2020) confirms a single regulatory framework to tackle a range of online harms, including disinformation. An expert group will build understanding and drive action to address disinformation. [5.5, paragraph 34] An Online Harms Bill to enact this policy is due before Parliament during 2021.

4.2 Alerting security and intelligence services to state-sponsored disinformation

Closely related to the DCMS Select Committee Inquiry and forthcoming Online Harms Bill are growing concerns about state-sponsored disinformation on social media, in particular Russian interference in Western democratic processes. In this context, the City research on ‘bots’ has been of interest to the UK Parliament’s Intelligence and Security Committee (ISC) and agencies worldwide responsible for national and international security and can be considered as a separate and further area of impact.

The ISC oversees UK intelligence and security activities and in November 2017 announced its inquiry into Russian activity against the UK. Ahead of a House of Commons debate on “Russian

Interference in UK politics” (21 December 2017), two official briefings confirmed the focus of the ISC inquiry as “the more than 13,000 Twitterbot accounts that were active during the referendum campaign and were deactivated after the ballot” and referenced (3.1) in three places. [5.6a, p. 2-3; 5.6b, p. 17, p. 20-21] City’s research was cited in the Commons debate, with one MP describing a “13,500-strong Twitter bot army”. [5.6c] The continuing focus on Twitterbots was confirmed with City’s research (3.1) directly referenced in a House of Commons Briefing Paper (26 March 2018) on “National Security and Russia” [5.7, p. 8] following the attempted murder of former double agent Sergei Skripal and his daughter in Salisbury earlier that month. When the ISC Committee’s Russia report was finally published in July 2020 it cited extensively the DCMS Inquiry, referring to Russia’s promotion of disinformation and attempts at broader political influence overseas, including significant bot activity on social media identified by “open source studies” [5.8, paragraph 28], and supported the DCMS Inquiry’s conclusion that “the UK is clearly vulnerable to covert digital influence campaigns” [5.8, paragraph 123]

Based on their Twitterbot research, Bastos and Mercea were invited in 2017/18 to several private / confidential meetings with security agencies to explain their research insights on how disinformation campaigns by state-sponsored actors and their impact can be tracked and evaluated. This included a joint Royal United Services Institute for Defence Studies / UK Cabinet Office expert round-table discussion on the impact of Russia’s influencing operations in the UK, meetings with NATO’s Strategic Communications Centre of Excellence and the UK Foreign & Commonwealth Office. In the public domain, in November 2017, [REDACTED] Canadian Security Intelligence Service (CSIS) unclassified seminar: “The Security Challenges of Modern Disinformation”. Part of CSIS’s academic outreach programme, the seminar brought together a multi-disciplinary group of experts from Canada, USA, and Europe. The seminar was held under Chatham House rules and examined the strategic impact of disinformation on national security and the integrity of democratic institutions. A report entitled *Who Said What? The Security Challenges of Modern Disinformation* [5.9] was based on the expert contributions and states baldly: “Disinformation poisons public debate and is a threat to democracy”. [5.9, p. 11] The report includes a detailed case study on Brexit and the “rise and fall of a Twitter botnet” [5.9, pp 51-58], clearly based on City’s research (3.1). The report calls for “Raised public awareness... to distinguish the real from the false” [5.9, p. 11]

4.3 Initiating public debate on the role of social media networks and companies in dis/misinformation and the use of ‘bots’ in their dissemination

A third key area of impact relates to how City’s discovery of the Brexit Twitterbot network led to a cascade of media attention putting pressure on Twitter to acknowledge this in public and respond to the DCMS Select Committee regarding the research [5.2, p. 78, Question 484]. Between October and December 2017, 220 national and international media outlets (print and broadcast) reported on the City research (3.1), including two articles in The Times on 31 October, which called for a change in UK law [5.10a], and on 9 November which questioned the government’s approach to investigating fake news and Russian interference. [5.10b] The high profile of the City research both in Parliament and international media prompted Twitter UK to respond to Damian Collins MP on 19 January 2018 [5.2, p.85] confirming that a sizeable proportion of the botnet accounts identified by City were deactivated by the company because their activity was in breach of its content and spam policies.

Separately, Twitter asked to meet with Bastos and Mercea in late 2017. The meeting is subject to a Twitter-imposed non-disclosure agreement, but the company invited City to apply to its research programme for funding from a new scheme they had instigated into election integrity. Bastos and Mercea successfully applied to Twitter for £90,000 for research on “the Brexit Value Space and the Geography of Online Echo Chambers” (1 April 2018-31 December 2021)

5. Sources to corroborate the impact (indicative maximum of 10 references)

5.1 Ball, James (20 October 2017) A Suspected Network Of 13,000 Twitter Bots Pumped Out Pro-Brexit Messages In The Run-Up To The EU Vote. BuzzFeed News. Available at: <https://www.buzzfeed.com/jamesball/a-suspected-network-of-13000-twitter-bots-pumped-out-pro>

(Accessed 5.3.21)

- 5.2** Digital, Culture, Media and Sports Committee (24 July 2018). *Disinformation and 'fake news': Interim Report*. House of Commons. Available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/363/363.pdf>
- 5.3** Digital, Culture, Media and Sports Committee (14 February 2019). *Disinformation and 'fake news': Final Report*. House of Commons. Available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/1791/1791.pdf>
- 5.4** Department for Digital, Culture, Media and Sports and Home Office (December 2020) *Online Harms White Paper*. Available at: <https://www.gov.uk/government/consultations/online-harms-white-paper>
- 5.5** Department for Digital, Culture, Media and Sports and Home Office (December 2020) *Online Harms White Paper: Full Government Response to the consultation*. Available at: <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper>
- 5.6a** House of Commons. *Russian interference in UK politics and society*, Library Debate Pack, 19 December 2017. <https://commonslibrary.parliament.uk/research-briefings/cdp-2017-0255/>
- 5.6b** House of Commons. *Russia 2017*. Library Research Briefing, 20 December 2017. Available at: <https://commonslibrary.parliament.uk/research-briefings/cbp-8157/>
- 5.6c** House of Commons. (2017) 21 December Debate. Volume 633. Available at: <https://hansard.parliament.uk/commons/2017-12-21/debates/9AF1EE0B-DE51-48A9-8633-D7DE3B7EFF96/RussianInterferenceInUKPolitics>
- 5.7** House of Commons (2018) *National Security and Russia*. Research Briefing, 26 March. Available at: <https://commonslibrary.parliament.uk/research-briefings/cbp-8271/>
- 5.8** Intelligence and Security Committee of Parliament (2020) *Russia*. Report, 21 July. https://isc.independent.gov.uk/wp-content/uploads/2021/03/CCS207_CCS0221966010-001_Russia-Report-v02-Web_Accessible.pdf
- 5.9** Canadian Security Intelligence Service (2018). *Who Said What? The Security Challenges of Modern Disinformation*. Report of seminar, Ottawa, 20 November 2017. Available at: http://publications.gc.ca/collections/collection_2018/scrs-csis/PS73-1-2018-02-01-eng.pdf
- 5.10a** Rifkind, H. (2017, 31 October). Fakers have a free rein over political adverts. The Times. Available at (paywall): <https://www.thetimes.co.uk/article/fakers-have-a-free-rein-over-political-adverts-r56qbw62>
- 5.10b** Rifkind, H. (2017, 9 November). Is Vladimir Putin meddling in British politics? The Times. Available at (paywall): <https://www.thetimes.co.uk/article/the-fight-against-fake-news-p007rjshk>