Martin Innes, Diyana Dobreva & Helen Innes (2019) 'Disinformation and digital influencing after terrorism: spoofing, truthing and social proofing,' **Contemporary Social Science**,

DOI: 10.1080/21582041.2019.1569714

Abstract

This article explores how digital communications platforms are used in the aftermath of terrorist attacks to amplify or constrain the wider social impacts and consequences of politically motivated violence. Informed by empirical data collected by monitoring social media platforms following four terrorist attacks in the UK in 2017, the analysis focusses on the role of 'soft facts' (rumours / conspiracy theories / fake news / propaganda) in influencing public understandings and definitions of the situation. Specifically, it identifies three digital influence engineering techniques – spoofing, truthing and social proofing – that are associated with the communication of misinformation and disinformation. After configuring these concepts, the authors consider their implications for policy and practice development, concluding that, to date, possibilities for evidence-informed post-event preventative interventions have been relatively neglected in the formulation of counter-terrorism strategies. They recommend more attention be paid to how strategic communications interventions can counteract the effects of misinformation and disinformation, and thus mitigate the wider public harms induced by terror events.

Keywords: Terror; Prevent Strategy; social media; disinformation; rumours; social reactions.

Introduction

> Technological environments are not merely passive containers of people but are active processes that reshape people and other technologies alike. ...the technology of electric circuitry represents one of the major shifts of all historical time. (McLuhan, 1962, p. 7)

Cast as a prophet of the information age, Marshall McLuhan's prescience was in identifying the complex and recursive influences that human behaviours and new technologies have upon each other. His insight represents a valuable counterpoint to the somewhat utopian and technologically deterministic 'early adopter' narratives that cast social media as an unparalleled public good, and a benign instrument of democratisation and freedom. In the wake of a number of public scandals in recent years involving these technologies, there has been a growing appreciation of the complex interactions between social and technological factors that are engaged by these modalities of communication and, as a consequence, how they can induce malign effects. To a significant extent, this shift reflects mounting political and public consternation about social media communications being harnessed to disseminate misinformation and disinformation routinely, and relatively invisibly, to influence collective behaviour and public attitudes across multiple situations and settings (Allcott & Gentzkow, 2017; Howard & Kollanyi, 2016).

Numerous studies attest to the digital information environment being awash with rumours, conspiracy theories and 'fake news' (Gonzalez-Bailon, 2017; Greenhill & Oppenheim, 2017;

Oh, Agrawal, & Rao, 2013) that travel further, faster and deeper than 'truths' (Vosoughi, Roy, & Aral, 2018). These informational forms are especially influential on public attitudes and behaviours in moments of emergency and crisis, such as terrorist attacks (Roberts et al., 2018).

Amoore & Piatukh (2015) contend that data processing algorithms are functioning as important 'instruments of perception', subtly framing which social issues are collectively attended to and neglected. This ability to steer the constructing of 'public problems' is pivotal to an understanding of how social communications platforms and their data are influencing social order (Couldry & Hepp, 2017; Gonzalez-Bailon, 2017). Margetts et al. (2016) have sought to deconstruct the roles of social media communications and networks in political mobilisations around such problems, concluding that, while social media technologies possess specific capacities to influence collective thought and action, they are fundamentally better at confirming and co-ordinating minds than changing them.

This article seeks to develop an understanding about how digital communications are used to influence the ways publics think, feel and behave during and after terrorist events. Specifically, it adduces evidence of how misinformation and disinformation that proliferates in the wake of terror events alters the social impacts of the violence. Accordingly, informed by empirical data generated through systematic monitoring of social media in the aftermath of four terror attacks that took place in the UK in 2017, the article conceptually delineates three specific techniques of digital influence engineering. 'Spoofing' involves trickery, deception or misdirection to misrepresent the identity of sources and/or the validity and reliability of information. These techniques are labelled 'identity spoofing' and 'information spoofing' respectively. 'Truthing' engages truth claims, including conspiratorial hidden

truths, or presenting factual information to try and persuade. ‘Social proofing’ utilises feedback mechanisms designed into social media platforms, comments, and bots, to construct an aura (sometimes illusion) of social support, on the understanding that it will influence the behaviour of other users.

Our focus upon reactions to terrorism is warranted because the periods following terror events are frequently moments of profound uncertainty, fraught with high social and political drama, where public sense-making processes are engaged to try and understand what has happened and why. As such, they occasion multiple attempts to persuade and influence, and social media platforms are increasingly integral to how individual and collective reactions unfold. Positioned in this way, the article asserts three principal claims:

1. Empirically, the analysis provides new evidence about reaction patterns to terrorism, and the roles played by misinformation and disinformation in shaping them.
2. The analysis demonstrates how social media analytics can generate evidence illuminating hitherto difficult to observe aspects of social life: in this case the temporal dynamics of social reactions to terrorist atrocities.
3. For policy and practice, the analysis affords an opportunity to think about how to manage reactions to terrorism, so as to minimise social impacts and harm and, more generally, about what works in constraining the influence on public perceptions and opinions of the kinds of ‘soft fact’ that seemingly thrive in the new media ecology.

The next section describes the research design and data, and key features of the four terrorist attacks. This context setting is important in understanding how and why varieties of soft fact circulate so well in the aftermath of a significant incident. Indeed, a central claim is that spoofing, truthing and social proofing behaviours are more likely to occur in a crisis tempo,

than during normal time. We define a soft fact as malleable and plastic information that may shape peoples' thoughts, feelings and behaviours, even if they lack the stability and authority of something ascribed the status of a hard fact (Innes, 2014). In effect, soft fact is a conceptualisation that seeks to draw out the family resemblances between a range of distinct, but related notions, including rumours, conspiracy theories, propaganda and fake news (Allport & Postman, 1947; Fine, Campion-Vincent, & Heath, 2005; Shibutani, 1966). Subsequent sections of the article detail each of the three main influencing techniques and their roles in communicating soft facts. The conclusion reflects upon the implications for policy and practice. It considers, specifically, how post-event prevention has been a relatively neglected feature of counter-terrorism strategy development, and the importance of managing misinformation and disinformation in the wake of an attack. More generally, the evidence and insights presented enhance our understandings of the workings of techniques of digital influence engineering, the ways these techniques are shaping the ordering of social reality, and how managing the resultant influence effects needs to be integrated within future public policy developments.

Data and method

The data reported here derive from a wider research programme, comprising several distinct, but linked studies, investigating social reactions to terrorist events. The programme has involved an extensive effort to collect, process and analyse social media data on the grounds that its streaming quality provides unparalleled opportunities to track and trace how public perceptions and behaviour unfold, thus affording a perspective unavailable via more orthodox social research methodologies.

The data were collected in the wake of four terrorist attacks that took place in the UK in 2017. The first attack on Westminster Bridge (22 March 2017), near the Houses of Parliament, was committed by a lone offender driving a van into pedestrians, before using a knife to fatally stab a police officer. This attack methodology was very different from the Manchester bombing of the Ariana Grande concert (22 May 2017), which was far more sophisticated in its planning and preparation. The Manchester attack was followed on 3 June 2017, back in London, by a marauding attack by three individuals, again using vans and knives. The fourth attack was committed a few days later by a single perpetrator targeting Muslim worshippers in Finsbury Park (19 June 2017).

A total of just over 30 million data points were collated from across multiple social media platforms utilising the Sentinel platform. Herein we focus principally on Twitter data, augmented by Facebook materials, anonymising data cited to reflect their often contentious contents. Sentinel comprises a suite of data collection and analysis 'apps' and algorithms, with similar collection and processing functionality to many commercial packages (Preece et al., 2018).Whereas these data are 'black boxed' (Pasquale, 2015), Sentinel is designed as a 'glass box', enabling researchers to investigate how manipulating particular decisions and choices in terms of data collection, processing and analysis, structure and shape the resultant data flows. Sentinel's data collection is organised around a series of 'channels', comprising around 400 search terms that can be configured by researchers in near real time, acting as filters, screening out irrelevant material and capturing units of social media traffic, which, because of their linguistic content, are likely to be connected to the subject of interest. This structure enables the system to work within the 1% limit of total traffic volumes that Twitter make freely available through 'the firehose' (the unfiltered full stream of tweets and their metadata).

Data analysis was driven by the conceptual interest in the role of soft facts in influencing how terror attacks are collectively interpreted by the public and assigned meaning. Earlier generations of scholars tended to study these information forms in isolation. However, because of the conditions pertaining to the contemporary media ecology, they are now routinely interpolated and overlapping. This quality needs to be articulated in how we study them today.

Reflecting the aim of describing and documenting key digital influence engineering techniques, data reduction was accomplished through the identification of a series of episodes by the researchers that appeared especially interesting and relevant. This was done both in real time as the event was unfolding through live monitoring of developments, and retrospectively once data collection had been completed. Episodes can be understood as defined events within the larger narrative that can be isolated and studied intensively to draw out wider learning in terms of what happens and why. In this sense, clear analogies can be made with the principles of Manning's (2016) 'pattern elaborative theory'. He suggests that, an interplay between 'exemplary evidence' and key theoretical precepts, can distil regularities and patterns in behaviour and conduct not previously recognised or perceived, which can be subject to subsequent empirical testing.

A total of 22 episodes involving the communication of one or more soft facts across the four attacks were identified for detailed case study analysis. Data associated with each episode were subject to qualitative analysis, including of text and imagery as appropriate. Throughout the paper we have anonymised quoted social media materials, except where the account identities were known to be deliberately falsified. In assessing the episodes, emphasis was

placed on analysing digital behaviours and how people 'do things to information, to do things with information', rather than just interrogating the contents of what they communicate. The analysis was directed towards understanding the proximate causes and consequences of specific soft facts, and their role in influencing how public understandings and definitions of the situation were constructed and contested. Each episode illuminated some particular facet of how soft facts function that, when blended together, enable a more comprehensive set of insights. The observed patterns distilled from this process informed the development of the three key concepts that are the focus of this article.

Measuring terrorisms harms

Within the literature on terrorism a small, but growing, cluster of studies attend to the challenges of empirically measuring the impacts and harms of terror events (English, 2015). Systematic empirical study of social reactions to major crime and terror events has hitherto been inhibited by methodological constraints. Such incidents routinely induce complex blends of behavioural, cognitive and affective effects that are differently distributed across particular audience segments, making it challenging, using orthodox social research methodologies, to isolate and connect proximate causes with specific impacts. The streaming quality of social media affords a potential to overcome these limitations since they permit connecting of events to reactions in ways enabling their study at scale and over time. That said, in terms of measuring influence, there has been an over-reliance on easily quantifiable 'reach' and 'impression' metrics (Gonzalez-Bailon, 2017; Margetts et al., 2016;).

Summarising what is known about public reactions to terrorism derived from more established social research methodologies such as opinion polling and semi-structured interviewing studies, Smelser (2007) highlights a tendency for intense but relatively short-

lived impacts, comprising: psychic numbing, involving a combination of disbelief, denial and suppression of affect; immediately followed by intense emotions of fear, anxiety, terror, rage and guilt; a surge in solidarity and scapegoating actions; and outpourings of sympathy.

Analysing reactions to the 9/11 terror attacks, Nacos, Block-Elkon & Shapiro (2011) contend that the prevalence and distribution of these responses is structured by demographic characteristics, especially gender and race. Moreover, they suggest that, the translation into political and social problems, can be longer lasting in their consequences than implied by Smelser (2007). Coherent with this line of reasoning, analysis by Oksanen and colleagues (2018) of the impacts of the November 2015 Paris terror attack, suggests that post-event fear 'travelled', with increases detected in Spain, Finland, Norway and the United States, as well as France. Along with the findings of several other studies, they further identified increases in hate crime (see also Roberts et al., 2018; Williams & Burnap, 2016).

Many of these impacts can be discerned in the four terror attacks that are the focus of this analysis. Comparing the three weeks after each incident with the same periods twelve months earlier, as represented in Figure 1, rates of hate crime reported to police increased for three out of the four attacks.

FIGURE 1 ABOUT HERE

Whilst some caution is required when interpreting these data, as they might be affected by increased public reporting or police recording of hate offences, read as an indicator of post-event harms, they are insightful in charting both a general pattern of increases in hate violence after terror attacks and important specific differences between incidents. Increases

do not inevitably occur, as evidenced by the Finsbury Park data, which was the only event instigated by a far-right extremist. It is within this complex and febrile environment, that we need to make sense of the kinds of misinformation and disinformation soft fact that are the focus of this analysis since, in such an atmosphere, they can increase social tensions, amplifying the overarching harm induced by a terror event.

Spoofing

In his discussion of algorithmic functions in high frequency financial trading markets, Mackenzie (2018) describes a number of attempts to spoof the algorithms to leverage competitive advantage and profit. Albeit focussed on machines rather than humans, aspects of his analysis are intriguingly redolent of Goffman's (1967, 1983) detailed dissections of how people, in their co-present encounters and interactions with each other, seek to misrepresent or mask aspects of their identities and/or motives. These interactional tactics can be an effort to hide discreditable aspects of their self to avoid embarrassment, and to smooth the transactions of the interaction to circumvent social awkwardness.

Our analysis suggests two master-types of spoofing are used in the aftermaths of terrorist attacks: identity spoofing and information spoofing. The former is where an individual claims to be someone or have a social status s/he does not really possess. Information spoofing involves misrepresenting the content of a message, through processes of falsification, suppression or amplification. Sometimes these techniques are blended together. Evidence can be found of spoofed (fake) accounts deliberately spreading soft facts following all four UK terror attacks.

Identity spoofing is commonplace in online environments. People routinely edit the versions of their selves that they project online and utilise multiple digital social identities for public performance (Marwick, 2013). Variations of this digital identity editing were evident across the dataset, but we focus on instances where misrepresentation or identity fabrication was especially pronounced. In particular, in the immediate aftermath of the Manchester Arena attack, a series of individuals used social media to communicate messages that falsely connected them to the incident. This is a behaviour we label 'victim-claiming'.

The key features of this phenomenon are exemplified by the following message:

> EVERYONE PLEASE RETWEET THIS HELP ME! THIS IS MY LITTLE BROTHER FRANK WE WENT TO THE CONCERT TONIGHT IN #MANCHESTER & NOW WE CANT FIND HIM PLS (RT 18,131) 22 May 2017, 1:55 AM

In the emotionally charged post-attack atmosphere, this single message was retweeted at least 18,131 times. It was entirely fabricated. The picture of the 'missing child' accompanying the text was originally published in an online clothes catalogue for children with Down's Syndrome. Rehearsing an aspect of the analysis that will be developed subsequently, the integration of visual images into these kinds of messages was used to validate the truth of the claim of victimhood. Similar communicative tactics were detected following the Westminster, London Bridge, and most prolifically, Manchester Arena attacks.

The social and psychological motivations for victim-claiming behaviour are complex and beyond the purview of this article. Fundamentally, it involves a particular version of identity spoofing, where the actors concerned are seeking to appropriate a social status associated with being directly afflicted by a major public event and involved in its public narration.

Although it is difficult to calibrate the impacts of such messages, they did appear to amplify the wider sense of harm attributable to the original incident.

A second form of identity spoofing was potentially even more malign. As part of the analysis, 47 accounts controlled by the Russian state were detected operating through social media trying to infiltrate and incite reactions among wider online communities following the four attacks. These accounts, operating influence and interference measures, were identified using open source material published by the US Congress, the Russian magazine RBC and the US media outlet NBC, that attributed just over 2,500 fake social media accounts to the Internet Research Agency in St Petersburg (Popken, 2018; Русяева & Захаров, 2017). Most of these spoofed accounts were not particularly influential, but eight were highly active and it is estimated their messages were viewed over 153,000 times. Significantly, these 'sock puppet' accounts mimicked identities and values distributed across the ideological spectrum, each speaking to defined online thought communities.

One of these spoofed identity accounts was @SouthLoneStar, who adopted a southern US, white, politically Republican persona, constructing messages accordingly. As evidenced in Figure 2, a clear anti-Islam agenda was being pursued, something explicitly developed in his messages.

FIGURE 2 ABOUT HERE

Across all four incidents, the Kremlin-backed sock puppet accounts repeatedly disseminated rumours, including about the identity of the perpetrator of the Westminster attack. At about 14.40 on 22 March 2017, rumours started circulating across social media platforms regarding

the identity of the attacker. These rumours asserted that the individual responsible for the death of five people was the infamous hate preacher Abu Izzadeen (aka Trevor Brooks). A few hours after this rumour first appeared, it was proven to be false. Izzadeen was in prison and could not have been the attacker. Its spread had been significantly accelerated and amplified by Channel 4 News repeating it at just after 19.00 that evening, subsequently triggering considerable opprobrium on social media.

Three sock puppet accounts tried to use the Izzadeen rumour to influence their followers. They elaborated on the news about Izzadeen by introducing an embellished narrative and associated ideologically loaded truth claims. All three projected a fairly extreme far-right agenda; their messaging focussed on Izzadeen's past as a hate preacher, and as a representative of the entire Muslim community:

> Here's suspect Abu Izzadeen urging British shoppers to convert to Islam, right on the street of #London!… https://t.co/DXEebwgZ0T (RT 2,015) 22 March 2017, 7:19 PM

> Abu Izzadeen represents all of the Muslim faith. Underdeveloped cult of hate and death ☪ #PrayForLondon… https://t.co/vPgoVMeK33" (RT 502) 22 March 2017, 7:33 PM

This small number of spoofed accounts was highly influential at least in terms of their reach, collectively generating 7,875 re-posts. Most influential was @TEN_GOP whose one post about the Izzadeen rumour generated a higher number of re-shares (N=3,958) than the first tweet broadcasting the breaking news. In this context, it is clear that, by hiding behind faked local identities, these accounts were attempting to spread divisive nudges among the audience.

Identity spoofing was frequently a predicate for information spoofing. For example, consistent with the 'Little Brother Frank' case study discussed above, 28 spoof appeals about missing children were documented following the Manchester attack and subsequently shown to be false. They all deployed well-documented techniques of social–psychological persuasion, including: establishing a sense of urgency and public call to action – 'please help me' – triggering an emotional and spontaneous audience response, without careful deliberation about source and credibility; inclusion of fabricated personal details and a back story, usually centring on familial relations, for example 'my little brother' or 'my best friend'.

Truthing

If spoofing works by falsification or misrepresentation, truthing persuades by claiming to be furnishing the audience with the real facts. One version involved the use of statistics, data, quotes and official statements to try and discredit other narratives. A second version relates to how proponents of conspiracy theories frame truth claims as part of their narratives, purporting to convey what really happened in relation to a contentious or contested episode. Importantly, when engaging in truthing behaviours, actors do not just undermine accounts, for example by labelling them 'fake news', but proffer a more or less plausible alternative.

Following the four attacks, multiple instances were found of ideological groups disseminating messages disrupting and discrediting the official police and government narrative. Designed to appeal to right-wing groups, one such episode pivoted around countering an official statement made by Mark Rowley (national lead for Counter Terrorism Policing and Acting Deputy Commissioner for the Metropolitan Police), by presenting the 'real truth' about the causes of the attack.

Following the Westminster attack, Rowley made two statements informing the public about what had happened and developments in the police investigation. In his evening press conference, he stated:

> …we must recognise now that our Muslim communities will feel anxious at this time given the past behaviour of the extreme right wing and we will continue to work with all community leaders in the coming days. (Rowley, 2017)

This particular comment, drawing attention to the potential for extreme right-wing violence, which constituted a small fraction of an otherwise lengthy, informative and reassuring statement, triggered a considerable number of negative reactions from supporters of far-right ideologies. These reactions escalated into the construction and dissemination of a meme. Two days after the attack, it was posted by several high-profile far-right groups and individuals, for example Tommy Robinson and the British National Party. It contained Rowley's image on the left, an extract of his far-right concerns quote on the right, and an alternative truth claim at the bottom: 'No mention of the concerns of the English community feeling anxious concerning Muslim terrorism and prime example of the liberalism that is killing England.'

Variants of this base narrative were repeated and elaborated, designed to resonate with people with similar values and world views, rather than appeal to a wider audience. The claim pivoted around the idea that the real cause of the terrorist violence was not just the activities of Islamists, but also a failure of state institutions to protect English / British communities:

> Typical liberal nonsense! How about the British community's [sic] under threat from Islamic terrorists? Stop appeasing and start acting. (RT 132) 24 March 2017, 12:24 PM

> Fuck them and fuck you Mr Rowley! What about us …… THE BRITS? The people you promised to protect and serve!! (RT 0) 24 March 2017, 9:44 AM

A total of 80 original far-right tweets specifically referring to Rowley's statement were detected, generating 2,082 re-tweets. This quite aggressive form of reactance evidences the point rehearsed earlier from the work of Margetts and colleagues (2016) about social media coordinating and confirming minds, rather than changing them. Rowley's statement was never going to alter the opinions and values of those he was speaking out against, and indeed he became akin to a 'condensation symbol' (a concept from Edelman, 1985) for their vitriol and shared concerns. Equally importantly, this episode conveys how, consistent with the predicates of behavioural science more generally, attempts to deliberately influence individual and collective conduct tend to acquire more traction when they go with the grain of how people are inclined to act.

Similar truthing techniques are routinely mobilised by conspiracy theorists (Hofstadter, 1964; Keeley, 1999). Multiple online groups recurrently construct false flag narratives around high-profile events, by selecting and connecting seemingly unrelated elements and details from the story, and presenting them as evidence and facts. Specific examples of these approaches identified across the four attacks included: a lack of bomb smoke or dead bodies on pictures published by media (Manchester); lack of blood on footage and lack of CCTV cameras in key areas (Westminster); contradictory witness reports (Finsbury Park); claims that no victims were ever taken to the hospital (London Bridge). Other conspiratorial authors asserted these events were 'psyops' (psychological operations) or false flags, with the people at the scene, including the victims, the police, and the medical personnel, cast as 'crisis actors' (trained

actors, role players, or volunteers) performing as part of a bigger conspiracy. Whilst conspiracy theories constituted a much smaller proportion of the total social media communications than the rumours regarding the missing children and Abu Izzadeen, they are an important part of the communications ecosystem in the immediate aftermath of terrorist attacks.

Truthing and spoofing techniques are often blended together. For example, highly publicised claims were made by the Facebook user Paula Robinson immediately following the Manchester bombing that she had personally escorted a large number of misplaced children to a nearby hotel and was looking after them until they were picked up by police. The first in a series of Facebook messages stated: 'Bomb gone off Manchester git loads with kids with us please pass on we taking them to premier inn bomb at Victoria station.' 23 March 2017, 12:30 AM [approximate time].

At the time, several truthing techniques endowed a surface validity to her story in that she gave: her real name and phone number; directly addressed worried 'parents'; shared a meme with her message; and, gave details as if she were at the scene. However, the following day both the hotel and Metropolitan Police confirmed that no 'missing children' were ever there.

Social proofing

In their recent analysis of the role of digital technologies in political mobilisation processes, Margetts and colleagues (2016) demonstrate how specific feedback mechanisms designed into a number of social media platforms function to shape users' collective behaviour. These social-psychological mechanisms refract the more general evidence from psychology about the role of social proof in processes of persuasion and influence, whereby people look to

others' experience as a reference point for their own conduct (Cialdini, 2009; Sharot, 2017). Transposed into an online environment, Anspach (2017) reports evidence that social proofing heuristics can influence social media user's selection, consumption and sharing of articles on Facebook. In particular, the illusion of large numbers can be highly deceptive, leading people to grossly overestimate how many others share their views, or how far they differ from the mainstream.

Returning to the episode involving Assistant Commissioner Rowley's statement outlined earlier, we can see how some individuals sought to harness social proofing mechanisms to their advantage. The former leader of the English Defence League posted a response message on Twitter under his public alias of Tommy Robinson: 'You treasonous coward'. Sympathisers sought to demonstrate their support for his statement. It received 1,081 retweets and 1,668 likes, accompanied by repeated personal insults and attacks directed towards Rowley (n=55). It is worth noting that although these are not especially large numbers of re-shares, they are nonetheless salient because they express support for a highly personalised attack on a senior police officer, during a significant national security event.The retweets contrasted markedly with comments logged on the Metropolitan Police Service's Facebook page, where the majority expressed sympathy and condolences for the death of PC Keith Palmer and praised the work of the police and emergency services. The concentration of one-sided views in large numbers, visible in the Twitter thread started by Tommy Robinson, demonstrated a deliberate effort to manipulate the tunnelling effect of social media, constructing an illusion of consensus and support.

Similar effects can also be simulated by automated accounts that manipulate markers of social validation, popularly known as 'bots'. Social media likes, re-shares and favouriting are

infused with this notion that social proof or validation is an indicator of the value or accuracy of a communication (Confessor et al., 2018; Hern, 2017). Some groups and individuals use technological solutions to enhance their social information, for example employing 'spambots' and 'botnets' to flood channels with messages that interpret events in ways consistent with their norms and values. Fundamentally, they amplify the public visibility of particular ideas implying considerable social support for these positions.

In relation to the rumour about Abu Izzadeen referenced previously, the research team identified 20 bot accounts tweeting identical content with a time difference of a couple of seconds:

> FergusMcPop: Sources say London attack was an Islamist attack carried out by Abu Izzadeen https://t.co/M90tVimwQP #London #PrayForLondon #LondonAttack (RT 0) 22 March 2017, 7:22 PM

> Mcpopg: Sources say London attack was an Islamist attack carried out by Abu Izzadeen https://t.co/M90tVimwQP #London #PrayForLondon #LondonAttack (RT 0) 22 March 2017, 7:22 PM

This cluster of accounts was presented as emanating from different individuals with different names, profile pictures and descriptions engaging in a form of identity spoofing. However, content analysis revealed identical Twitter feeds for all accounts. In addition, quantitative characteristics showed almost simultaneous messaging patterns. The purpose of these bots seemed to be driving traffic to a particular website (gpn100.com) that no longer exists. Even though these spambots did not generate any considerable traction among the wider public, they did increase the volume of misinformation circulating at an important moment.

Evidence and insights for counter-terror policy and practice

The preceding discussion has highlighted how new social communication technologies have induced complex processes of public sense-making in the wake of political violence. The empirical thrust of this analysis has been to argue that blends of spoofing, truthing and social proofing are being implemented by a range of actors, to disseminate soft facts that influence public understandings and definitions of the situation, and ultimately the level of harm generated by terrorist violence.

Social media complexify how public understandings are produced in the aftermath of major events, but they also render these construction processes observable. Important evidence can therefore be distilled from these sources about the social organisation of public reactions to terrorist violence. Through careful analysis, past events can teach us about reaction patterns, enabling learning about what kinds of intervention amplify or constrain social impacts. Ultimately, the harm and impact of the violence are not solely inherent in the incident itself, but depend, in part, upon the dynamics of social reaction. As such, it can be manipulated via the kinds of digital influence engineering procedures illuminated in the preceding passages.

This observation offers potentially important insights for counter-terrorism policy. Specifically, considerable emphasis has been placed on the value of preventative and pre-emptive interventions, for example through the 'Prevent' component of the UK government's CONTEST strategy (HM Government, 2011; Zedner, 2007). The latter is the UK's cross-government approach to combatting international terrorism, within which, the Prevent strand is focussed on inhibiting and interdicting processes of violent radicalisation prior to

individuals becoming involved in violence. By contrast, this study's research evidence and insights suggest a need for policy and practice to attend more concertedly to the dynamics and mechanics of post-event interpretation and sense-making. This would amount to a 'post-event Prevent' approach that would establish preventative operating procedures designed to inhibit and mitigate attempts to manipulate the harm induced.

Aspects of what a post-event Prevent approach might look like are hinted at by examples in the dataset. Notably, the statement made by Assistant Commissioner Rowley, where he exposed 'the past behaviour of the extreme right wing' anticipating they would commit hate crime and stoke social division. His pre-emptive intervention represented an important innovation in the police's strategic communications, a shift triggered by research evidence derived from the terrorist murder of Lee Rigby in 2013, set out, for example, by Roberts and colleagues (2018).

Not all police communications strategies were equally successful. For instance, within twenty minutes after the Manchester Arena terrorist attack, Greater Manchester Police, quickly communicated via Twitter that they were investigating an incident in the City Centre, asking the public to avoid the area. A series of social media posts from the police continued to advise the public to stay away and signposted them to official updates from their Twitter account @gmpolice. However, aside from confirmation of the numbers involved an hour later, police tweets over the next four hours did little more than put the public in a holding pattern of 'stay-away' and 'working-at-the-scene' messages, that did not appease growing public frustration for more information. As a result, the information vacuum in a chaotic and rapidly developing situation became receptive to influencing behaviours by other actors.

Framed by such insights and the evidence on which they are based, the intent underpinning this analysis has been to develop a more nuanced understanding of the conduct of digital influence engineering. Delineating the techniques of spoofing, truthing and social proofing starts to promote such understanding, but it also raises further important issues: one being a need to develop richer and more nuanced ways of empirically measuring the actual social influence of such communications. For instance, considerable attention has been paid to the role of bots spreading rumours and other forms of disinformation. However, it does not confirm what actual influence, if any, these communications have upon real peoples' perceptions or opinions, especially as public awareness about their presence increases.

Developing better metrics for digital influencing is important in policy terms to overcome the risk that energy for reform gets captured by an intriguing, but not especially consequential feature of the new media ecosystem, because it is relatively visible. Our work shows that policy innovations have not focussed sufficiently on other more hidden persuaders, such as the discovery that foreign political actors are deliberately seeking to use social media communications to excacerbate social tensions in the wake of the four terror events. That geo-political conflicts are being threaded through more locally situated ones, via the digital communications infrastructure, raises serious challenges for policy development that may well require profound shifts in Western governments' security postures.

Drawing back from the situated details of specific episodes, what can be distilled from this analysis is evidence of how communication of soft facts in the wake of politically motivated violence is used to try and amplify public harm by functioning as an accelerant for increasing social tension. Police and their governmental partners need to attend far more to the ways communications interventions delivered in a crisis moment can either reassure, or deepen, the

collective impacts. Terrorist violence is fundamentally intended to terrorise, polarise and mobilise different segments of its audience. Accordingly, understanding how to design and deliver interventions that interdict such processes, and enable counter-influences, is likely to be an increasingly important component in managing future terror attacks.

Conclusion

By early 2017, senior police and security officials had been warning for some time that it was unlikely they could prevent all terrorist plots. However, no-one anticipated the UK would experience four attacks in three months. Faced with an ongoing stream of threats of varying degrees of complexity and sophistication, it is improbable that all future attacks can be interdicted. It does though seem more plausible to suggest that governments could learn lessons from past atrocities, in terms of how to respond to reduce the impact of terrorist provocations. Yet, the issue of how best to manage the public impacts of terrorism remains relatively neglected when compared with the amount of attention and effort directed towards understanding processes of radicalisation (English, 2015). A post-event Prevent Strategy would harness the evidence and insights being distilled via social media analytics to construct a suite of interventions targeted towards managing the aftermaths of atrocities. Even if it is not possible to prevent all such harms, it is possible to be smarter and more effective in limiting their impacts.

Documenting and describing instances of spoofing, truthing and social proofing starts to map out how and why some disinforming soft facts influence public understandings of specific terror attacks. Collectively, these concepts point to the complex and contingent public sense-making process that has been occasioned by the presence of social media in the new media ecology, in respect of major public events (Couldry & Hepp, 2017). Where previous

generations would have been dependent on information channelled via mainstream media, now a far more complex set of information feeds has been created. Careful and rigorous study of these connects us to the deeper sociological truism that the societal consequences and impacts of these major public events are as much a function of how we react, as they are of the original violence.

References

Allcott, H. & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. doi: 10.3386/w23089

Allport, G. W. & Postman, L. (1947). *The psychology of rumor*. New York: Henry Holt & Co.

Amoore, L., & Piotukh, V. (2015). Life beyond big data governing with little analytics. *Economy and Society*, *44*(3), 341–366. doi: 10.1080/03085147.2015.1043793

Anspach, N.M. (2017). The new personal influence: How our Facebook friends influence the news we read. *Political Communication*, *34*(4), 590–606. doi: 10.1080/10584609.2017.1316329

Cialdini, R.B. (2009). *Influence: Science and practice*. New York: William Morrow.

Confessor, N., Dance, G.J.X, Harris, R., & Hansen, M. (January, 2018). The follower factory. *The New York Times*. Retrieved from https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

Couldry, N., & Hepp, A. (2017). *The mediated construction of reality*. Basingstoke: Palgrave Macmillan.

Edelman, M. (1985) *The symbolic uses of politics*. Chicago: University of Illinois Press.

English, R. (2015). *Does terrorism work?* Oxford: Oxford University Press.

Fine, G.A., Campion-Vincent, V., & Heath, C. (2005). *Rumor mills: The social impact of rumor and legend*. New Brunswick: Aldine/Transaction.

Greenhill, K.M., & Oppenheim, B. (2017). Rumor has it: The adoption of unverified information in conflict zones. *International Studies Quarterly*, *61*(3), 660–676. doi: 10.1093/isq/sqx015

Goffman, E. (1967). *Interaction ritual: Essays in face to face behaviour*. New York, NY: Pantheon.

Goffman, E. (1983). The interaction order. *American Sociological Review*, *48*(1), 1–17. doi: 10.2307/2095141

Gonzalez-Bailon, S. (2017). *Decoding the social world*. Cambridge, Ma.: MIT Press.

Hern, A. (June, 2017). Facebook and Twitter are being used to manipulate public opinion – report. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/jun/19/social-media-proganda-manipulating-public-opinion-bots-accounts-facebook-twitter

H.M. Government (2011). *Prevent Strategy*. London: Crown Stationery Office.

Hofstadter, R. (1964). *The paranoid style in American politics and other essays*. Cambridge: Harvard University Press.

Howard, P.N., & Kollanyi, B. (2016). *Bots, #Stronger, and #Brexit: Computational propaganda during the UK-EU referendum*. Retrieved from https://arxiv.org/abs/1606.06356.

Innes, M. (2014). *Signal crimes: Social reactions to crime, disorder and control*. Oxford: Oxford University Press.

Keeley, B. (1999). Of conspiracy theories. *The Journal of Philosophy*, *96*(3), 109–126. doi: 10.2139/ssrn.1084585

MacKenzie, D. (2018). Material signals: A historical sociology of high frequency trading, *American Journal of Sociology*. *123*(6), 1635–83.

Manning, P. (2016). Goffman and empirical research. *Symbolic Interaction*, *39*(1), 143–52. doi: 10.1002/SYMB.220

Margetts, H., John, P., Hale, S., & Yassera T. (2016). *Political turbulence: How social media shape collective action*. Princeton: Princeton University Press

Marwick, A. (2013). *Status update: Celebrity, publicity and branding in the social media age*. New Haven: Yale University Press.

McEnery, T., McGlashan, M. & Love, R. (2015). Press and media reaction to ideologically inspired murder: The case of Lee Rigby. *Discourse and Communication*, *9*(2), 1–23. doi: 10.1177/1750481314568545

McLuhan, M. (1962). *The Gutenberg galaxy: The making of typographic man*. Toronto: University of Toronto Press.

Nacos, B., Block-Elkon, Y. & Shapiro, R. (2011). *Selling fear: Counterterrorism, the media and public opinion*. Chicago: University of Chicago Press.

Nicas, J. & Frenkel, S. (February, 2018). Facebook and Google struggle to squelch 'crisis actor' posts. *The New York Times*. Retrieved from https://www.nytimes.com/2018/02/23/technology/trolls-step-ahead-facebook-youtube-florida-shooting.html

Oh, O., Agrawal, M., & Rao, H.R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly*, *37*(2), 407–426. doi: 10.25300/MISQ/2013/37.2.05

Oksanen, A., Kaakinen, M., Minkkinen, J., Räsänen, P., Enjolras, B., & Steen-Johnsen, K. (2018). Perceived societal fear and cyberhate after the November 2015 Paris terrorist attacks. *Terrorism and Political Violence*, 1–20. doi: 10.1080/09546553.2018.1442329

Pasquale, F. (2015). *The black box society*. Cambridge: Harvard University Press.

Popken, B. (February, 2018). Twitter deleted 200,000 Russian troll tweets. Read them here. *NBC News*. Retrieved from https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731

Preece, A., Spasić , I., Evans, K., Rogers, D., Webberley, W., Roberts, C. & Innes, M. (2018). Sentinel: A codesigned platform for semantic enrichment of social media streams.

118 IEEE Transactions on Computational Social Systems, *5*(1), 118–131. doi: 10.1109/TCSS.2017.2763684

Roberts, C., Innes, M., Preece, A. & Rogers, D. (2018). After Woolwich: Analysing open source communications to understand the interactive and multi-polar dynamics of the arc of conflict. *British Journal of Criminology*, *58*(2), 434–454. doi: 10.1093/bjc/azx024

Rowley, M. (2017). *Statements issued following attack in Westminster*. Retrieved from http://news.met.police.uk/news/latest-on-westminster-incident-229843

Sharot, T. (2017). *The influential mind*. New York: Little Brown.

Shibutani, T. (1966). *Improvised news: A sociological study of rumor*. Oxford: Bobbs-Merrill.

Smelser, N. (2007). *The faces of terrorism: Social and psychological dimensions*. Princeton: Princeton University Press.

Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. *Science*, *359*, 1146–1151. doi: 10.1126/science.aap9559

Williams, M.L. & Burnap, P. (2016). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, *56*(2), 1–28. doi: 10.1093/bjc/azv059

Zedner, L. (2007). Pre-crime and post-criminology. *Theoretical Criminology*, *11*(2), 261–81. doi: 10.1177/1362480607075851

Русяева, П., Захаров, А. (2017). *Расследование РБК: как «фабрика троллей» поработала на выборах в США* [RBC Investigation: how the 'troll factory' influenced the US elections]. Retrieved from https://www.rbc.ru/magazine/2017/11/59e0c17d9a79470e05a9e6c1

FIGURES

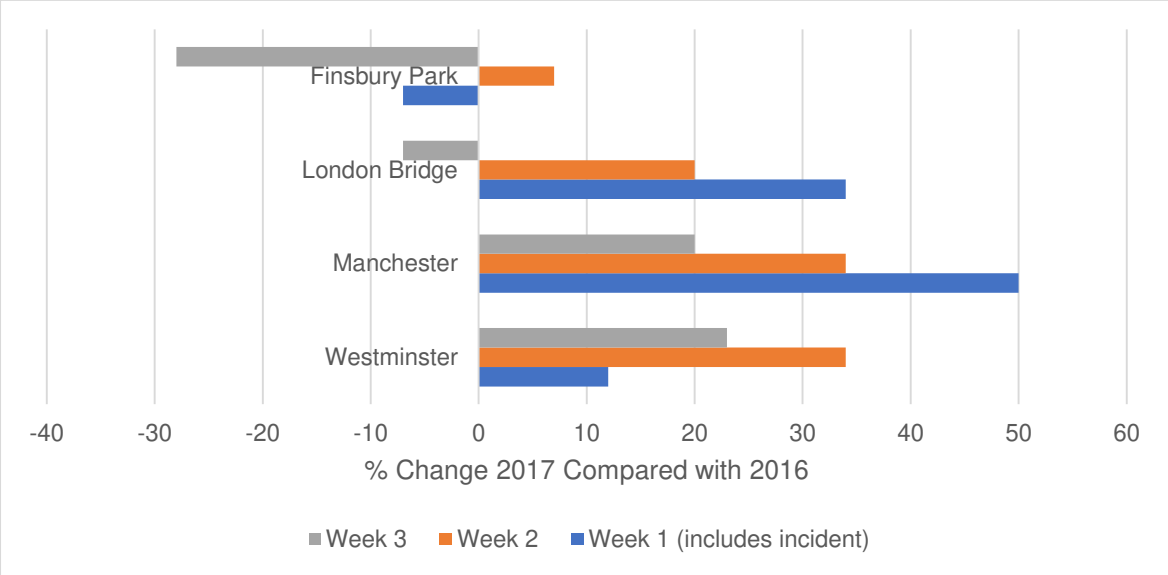

Figure 1: Percentage change in hate crimes in England and Wales following the 2017 terror attacks compared with the same period in 2016.

Source: Authors' own compilation from National Police Chief's Council Data.

Figure 2: Screenshot of a spoofed profile

Source: MashableUK, Retrieved from https://mashable.com/2017/11/14/troll-fake-muslim-picture-westminster-attack-russian-bot/?europe=true