Harms of digital capital: social harm analysis of online public resistance and information pollution

Cover page (not anonymised)

Harms of digital capital: Social harm analysis of online public resistance and information pollution

Author names and affiliations:

Anita Lavorgna
University of Bologna, Italy

Pamela Ugwudike
University of Southampton, UK

Yadira Sanchez-Benitez
University of Southampton, UK

Abstract:

Criminological studies of social harms extensively document intersections of power and the production of harm, revealing how the actions of the powerful in the public and private sectors expose (typically) less powerful groups to harm, often with impunity. Whilst this scholarship provides much needed insight into the often minimised or dismissed harms of the powerful, attention must also be paid to the agency of the victimised and the outcomes of their active efforts to resist such harms, especially in a digital context where concepts such as 'power' and 'capital' might take a different meaning. To this end, this paper expands existing criminological scholarship on social harms by providing new insights on how the dynamics of resistance by ordinary citizens, that is, people not generally considered part of the powerful capitalist elite, can nevertheless produce secondary social harms. The paper uses the example of online resistance to the COVID-19 digital tracing 'track and trace' app in England and Wales to unravel how ordinary citizens utilise their agency to resist the perceived harms of powerful actors whilst at the same time producing the secondary social harm of information pollution.

Key words/short phrases:

Word count:

10,104

Harms of digital capital: Social harm analysis of online public resistance and information pollution

Anita Lavorgna (University of Bologna), Pamela Ugwudike (University of Southampton, Yadira Sanchez-Benitez (University of Southampton)

Abstract

Criminological studies of social harms extensively document intersections of power and the production of harm, revealing how the actions of the powerful in the public and private sectors expose (typically) less powerful groups to harm, often with impunity. Whilst this scholarship provides much needed insight into the often minimised or dismissed harms of the powerful, attention must also be paid to the agency of the victimised and the outcomes of their active efforts to resist such harms, especially in a digital context where concepts such as 'power' and 'capital' might take a different meaning. To this end, this paper expands existing criminological scholarship on social harms by providing new insights on how the dynamics of resistance by ordinary citizens, that is, people not generally considered part of the powerful capitalist elite, can nevertheless produce secondary social harms. The paper uses the example of online resistance to the COVID-19 digital tracing 'track and trace' app in England and Wales to unravel how ordinary citizens utilise their agency to resist the perceived harms of powerful actors whilst at the same time producing the secondary social harm of information pollution.

Introduction

Both the extant criminological and zemiological scholarship on social harm provide much needed insight into the often minimised or dismissed harms of the powerful. However, attention must also be paid to the agency of the victimised and their active efforts to resist harms. On the importance of studying the agency of those victimised by social harms of the powerful, Hillyard and Tombs' (2021) position is somewhat instructive. They argue that 'if we are attempting to measure both the nature and the relative impact of harms which people bear, it is at least reasonable to take some account of people's own expressions, and perceptions, of what those harms are!' We respond to this call by drawing on large scale social media data to study both the agency of the victimised and their ability to produce secondary social harms when resisting perceived primary harms of the powerful.

The concept of 'resistance' has been traditionally invoked in defence of the oppressed or with reference to forms of civil disobedience. Over time, the notion of a 'right to resist' has gained currency with governments, civil society organisations, and other social institutions in liberal democracies (Cosi, 1984; Ferrari, 2004: 177). The notion conjures up the image of powerless masses rejecting the efforts of a powerful elite to control them. In this scenario, the masses are typically marginal individuals, and their resistance is often discussed with a focus on the powerful elite (that is, the efforts of the powerful to control dissent). Resistance can take the form of both action and speech (Krippendorff and Halabi, 2020), including digital speech (e.g., Ziccardi, 2012; Hill, 2018).

This paper focuses on online resistance and demonstrates that power relations can become more fluid in online spaces as new forms of agency become possible, empowering historically marginalised groups resisting the harms of the powerful to exercise power by producing discourses that can have adverse social, health and other implications. We, therefore, show that 'acts of resistance' perpetrated by the less powerful (that is, those who have suffered social harms because of unequal power structures, keeping in mind that social harms need to be considered in relational terms – see Canning and Tombs, 2021) can, at times, also create socially harmful implications. In doing so, we adopt a subversive stance that reflects the traditional criminological focus on the harms of the less powerful. A key difference between our approach and mainstream criminological accounts of the aetiology of harm is that we develop a structural outlook that focuses on the ways in which acts of resistance by the less powerful to primary harms of the more powerful can potentiate secondary harms in the form of profound adverse implications for individuals and communities.

In the paper, we draw on the empirical example of online resistance against the harms linked to a COVID-related public health directive introduced by the government and endorsed by other powerful groups including the media and politicians. Drawing on this example, we explore the agency of those involved in resisting perceived harms posed by the actions of the powerful national state in introducing and promoting the COVID-19 digital tracing 'track and trace' app (henceforth, 'the app'). We focus on resistance on Online Social Networking Sites (ONS), specifically Twitter. In doing so, we unravel how online social resistance by the less powerful to perceived primary harms of the more so can potentiate secondary harms. Through our analysis of these issues, we advance the criminological scholarship on social harms.

The criminological scholarship on social harm

The social harm approach emerged towards the end of the 1990s, in part to counter criminology's inordinate focus on the legal definition of crime as acts or omissions that violate the criminal law (Hillyard et al., 2004). Inspired by this reductionist legal definition, criminology predominantly focused on the activities of the less powerful: the group most likely to be criminalised because of power imbalances and intersectional oppressions (Canning and Tombs, 2021: 33). Rejecting this, the proponents of a new social harm perspective argued that moving criminological analysis beyond 'crime' to the broader concept of 'harm' was necessary to accommodate the socially harmful, but often non-criminalised, actions of more powerful groups. Such harmful acts have been defined by the original proponents of the approach as 'the deleterious activities of local and national states, and of corporations, upon people's lives (Hillyard and Tombs, 2017: 16).

Of course, the social harm approach was influenced by the seminal work of earlier scholars, *in primis* by the contributions of exponents of sociological criminology such as Ferri (1921) or of the Chicago school and particularly Sutherland (1983), but also by the work of figures such as Durkheim (1893; 1900), and others (e.g., Schwendinger and Schwendinger 1977). Together, regardless of the different focuses of their analyses, they all emphasised that, several socially harmful practices of the powerful are not criminalized, and criminology should look beyond the legalistic understanding of criminal justice systems[1]. Indeed, the social harm approach in criminology views the powerful and their activities as fitting objects of criminological enquiry not least because of the capacity of powerful institutions, groups and elites to produce harms that are typically minimised and deemed outside the purview of criminal justice intervention. The field has evolved in different theoretical directions, focusing on diverse definitions of social harm, its perpetrators, and its victims (e.g., Canning and Tombs, 2021; Greenfield and Paoli, 2013; Pemberton, 2015; Leighton et al., 2021), but one of the several themes unifying the multifarious strands of social harm scholarship is relevant to this paper. It concerns the study of non-criminalised practices that cause harm and should, as such, attract more criminological and policy attention than crime (Hillyard and Tombs, 2004).

Perhaps inspired by the perspective's initial emphasis on harms produced by those more powerful in society, criminological studies of social harms tend to focus on how the activities

[1] For further discussions on this point, see also Lasslett, 2010; Yar, 2012; Rothe and Kauzlarich, 2016.

of the more powerful, including corporations and nation states, produce harms, primarily victimising the less powerful in society (see, generally Canning and Tombs, 2021). Some emphasise the links between such harms and capitalist imperatives (e.g., Pemberton, 2015). Thus, scholars adopting the social harm approach have explored cultural, social, political, and other activities that produce harm, including harms related to the environment (e.g., White, 2021), employment (e.g., Tombs, 2014; Lloyd, 2021), industrial food production (e.g., Leighton, 2021) and penal practices (e.g., Presser, 2013).

That which is missing from the social harm literature is an understanding of how less powerful groups who feel victimised by the primary harms of the powerful actively resist such harms and how, at times, secondary harms can arise from such resistance. This paper addresses the gap in knowledge. By focusing on the potential secondary harms generated by ordinary citizens digitally resisting the harms of the powerful, the article advances the criminological scholarship on social harms. It moves beyond the emphasis of many criminological study of social harms on how the actions of the powerful expose (typically) less powerful groups to harm, often with impunity (Canning and Tombs, 2021). It also advances the scholarship on online social resistance by highlighting how resistance in digital spaces such as ONSs can produce socially harmful implications.

Criminological studies influenced by the social harm perspective have similarly explored digitally enabled harms of various kinds (e.g., Lavorgna 2021a; Gordon et al., 2022). However, the studies have not addressed the issue in the context of public resistance to perceived harms of the powerful. To shed some light on the nature of digitally enabled harms, Wood (2021) has conceptualised several harms emerging from technology design and deployment, of which the harm of zemiosis is the most relevant in this context. It refers to the harms produced by sociotechnical systems which rely on human interaction with technology. In his study of the 'zemiogenic power of technologies', Wood (2021) provides a detailed analysis of how new and emerging technologies propagate a range of harms. Key examples include the harms of algorithmic racial discrimination in justice systems (as detailed also in Ugwudike, 2020) and the disempowering practices of social media algorithms which can reproduce the offline marginalisation of less powerful groups in online spaces (see Lavorgna et al., 2021a). In our analysis of how secondary harms can emerge from resistance to harms of the powerful, we explore the digitally enabled harm of information pollution on ONSs and use it as an empirical and illustrative example.

The advent and proliferation of ONSs, such as Twitter, mean that forms of social, cultural and political expression now transcend spatiotemporal boundaries. With affordances provided by the sites, millions of people globally are connecting through such virtual platforms. Affordances in this context refer to the opportunities provided by social media platforms which empower anyone with a compatible device (e.g., a smart phone) to broadcast their opinions and resistance to a global audience in real time, changing the equilibria of key elements of interest in investigating social harms: that is, how language is used and exercised (Canning and Tombs, 2021: 117). The affordances are transforming the sites into global platforms of networked interaction, information dissemination, and forms of social resistance, such as expressions of dissent and opposition towards the perceived harms of the powerful (Bonilla and Rosa, 2015; Carney, 2016; Freelon et al., 2016; Lavorgna 2021b; Lavorgna et al., 2021a). Although they currently facilitate such exercise of public agency, this paper demonstrates that the social media affordances that enable public resistance can in themselves empower users, including those from historically marginalised groups, to engage in forms of resistance that propagate secondary harms. In this paper, we focus on the secondary harm of information pollution.

Information pollution as a secondary harm of resistance

The concept of information pollution was first used by Nielsen (2003) to describe irrelevant, redundant, unsolicited, and low-value information, and more recently by Wardle and Derakhshan (2017) to describe a distinguishing feature of political life in online spaces. It relates to the practice of constructing and disseminating information that can cause harm to others, and it can take the form of misinformation (when false information is shared, but no harm is intended), disinformation (when false information is knowingly shared to cause harm), and malinformation (when genuine information is shared, but harm can be caused[2]). The notion of information pollution has since been applied in socio-criminological analysis, with specific reference to the COVID-19 pandemic, and conceptualised as capable of producing social harms such as discouraging public compliance with public health initiatives (Lavorgna, 2021b; Lavorgna et al., 2021a). Looking at information pollution in its entirety allows us to overlook intent, which is often difficult to prove and can be irrelevant to the emergence of a certain harmful outcome.

[2] For the context of this contribution, the motive does not matter.

Conspiratorial thinking can also be classified as a dimension of information pollution. It manifests itself in the form of speculations about the existence of socially harmful conspiring agents such as public institutions, government representatives, and others (see also, Coday, 2006). Conspiratorial thinking is located at the basis of some of the epistemic mechanisms enabling the proliferation and success of information pollution (Lavorgna, 2021b). The existence of conspiratorial thinking in the context of the pandemic has been largely studied from other disciplinary standpoints (e.g., Biddlestone et al., 2020; Romer and Hall Jamieson, 2020; Allington et al., 2021; Pummerer et al., 2022). This increased attention to the topic is not surprising, considering that conspiratorial thinking thrives in situations where people's need to feel safe and secure in their world and to exert control over their existence are threatened. Such thinking helps restore feelings of agency and power (Bangerter et al., 2020; Douglas et al., 2020; Imhoff and Lamberty, 2020).

Information pollution is a rapidly growing problem and a type of harm that can be perpetrated by individuals/ordinary citizens against communities, as noted by the World Health Organisation (WHO) (2020) in their description of what they conceptualise as a fast growing 'infodemic'. Albeit it is a non-criminalised harm that has been largely ignored within social harm or zemiological scholarship when focusing on the harms of the powerful. Nevertheless, in this paper, we demonstrate how, in the process of resisting perceived (primary) harms of the powerful, online information pollution can emerge as a (secondary) socially harmful practice that warrants criminological and zemiological attention. In as much as social harm includes actions that produce physical harm (Hillyard and Tombs, 2017; Canning and Tombs, 2021) or harm the health of others (Tift and Sullivan, 2001), information pollution pertaining to public health initiatives can be categorised as a non-criminalised social media practice that can produce a variety of social harms and can, accordingly, be approached from a social harm perspective (see Lavorgna, 2021b). Indeed, studies suggest that information pollution via misinformation, disinformation, malinformation or conspiratorial thinking can be socially harmful because it can adversely influence people's beliefs, attitudes, and behaviours (Chadwick et al., 2018). Online pollution of health information, for instance, can generate potentially harmful attitudinal and behavioural resistance to important public health initiatives, often impacting those socially more vulnerable (Lavorgna, 2021b). Where certain communities (e.g., BAME groups) are affected, this can worsen existing health inequalities. Official statistics and other studies reveal that BAME communities experience the worst health

outcomes in several areas of healthcare delivery including public health responses to the current pandemic (e.g., ONS, 2020; Royal Society for Public Health, 2020).

The potential for information pollution to disproportionately harm certain communities, very much reaffirms its nature as a social harm (see also, Lavorgna and Di Ronco, 2019; Lavorgna 2021b). As Hillyard and Tomb (2004: 23) observe, the social harm approach should be more concerned with 'mass harms' and the implications for 'a harmed community' (19), comprising 'groups of people in some form of collectivity who are physically or financially harmed by whatever means' (Hillyard and Tombs, 2004: 19). Unlike much criminological analysis, the approach is not individualistic; its sole focus is not on 'an atomised individual' and the criminalisation of the behaviour of individuals. Such a focus excludes social harms that should not and cannot be condensed into current reductionist definitions of interpersonal crime.

That said, the social harm literature places greater emphasis on the harmful activities of non-individuals such as corporations and states. Nevertheless, the non-criminalised but very harmful activities individuals perpetrate against communities deserves similar attention. Examples include sexism, bullying, racially motivated microaggressions, and other harms perpetrated online, including on social media platforms. For affected communities, these can produce adverse psychological and other injuries which the proponents of social harm theory and research recognise as social harms.

We expand the social harm literature by analysing harms perpetrated by individuals resisting harms of the powerful and who are empowered by ONSs to propagate (as a form of online social resistance) socially harmful information pollution.

Methodology

Data collection

Our study drew on a subset of Twitter data produced by a research project that brought together criminological and computational expertise to unravel the sociotechnical dynamics of people's resistance to the National Health Service of the United Kingdom (NHS) COVID-19 tracing app across England and Wales. The data comprised a large dataset of tweets opposing the use of the NHS Covid tracing app covering the period March-December 2020, and collected

retrospectively using WebDataRA (for further details, see Lavorgna et al., 2021a, b). WebDataRA is a Chrome browser extension that monitors the pages a researcher browses (e.g., social media timelines and search results) and saves relevant data and metadata as a spreadsheet (Carr, 2020).

The subset of data on which this paper focuses comprises Twitter accounts belonging to ordinary citizens,[3] or in other words, 'less powerful' actors compared with the typical (powerful) subjects of social harm research: corporations, states, and other powerful entities. To select the subset from the large Excel file of Twitter data, we used Excel filters which identified accounts that had between 200 and 1,000 followers. Out of these, we deleted the accounts that were considered 'dormant' accounts, that is, social media accounts, human or bot-operated, that have not posted or engaged with other accounts for an extended period of time (Wardle, 2018) – in our case, accounts that have tweeted on any subject less than once a week. While there is no agreed definition of 'dormancy' in relation to time, one week was considered a suitable time to show lack of engagement – see Lavorgna et al., 2021a, b. We also deleted accounts that appeared to be automated (bots) because they tweeted more than 50 times per day (a standard approach to detect social bots on Twitter, see Alothali et al., 2018). As our focus was limited to England and Wales, we excluded accounts with a different geographical location. Further, we filtered the tweets associated with the selected accounts by considering only those containing the keyword 'track and trace', to ensure that we were focusing on those directly engaging in conversations around the topic. These processes reduced our dataset to 7,496 profiles and 10,175 tweets. We then applied the Twitter Engagement metric, which is calculated based on the number of favourites, retweets and mentions generated by tweets, and measures the effectiveness of tweets and how much certain profiles are connecting with their followers (see also, Agarwal et al., 2019). The Twitter Engagement metric revealed that a total of 3,127 tweets belonging to 2,021 accounts were above average and were as such those with which users most actively engaged. These were analysed for this study as explained in the following section.

[3] 'Ordinary citizens' are generally defined, in negative terms, as 'non-elite' citizens, see, e.g., Peter and Zerback, 2020; 'elite' are those 'with power' as they are 'with vastly disproportionate control over or access to a resource', in line with the traditional sociological definition – see Khan, 2012. For the sake of our study, we consider 'ordinary citizens' to encompass both 'private citizens' representing themselves only, and citizens speaking on behalf of a grassroots initiative organized by a group of ordinary citizens (in line with Hopmann and Shehata, 2011).

It is important to note that, in order to carry out our study, we took all the necessary precautions to ensure that our research was ethically-informed (Zimmer and Kinder-Kurlanda, 2017): we collected data by using a software tool compatible with Twitter policies, and we collected information that had already been posted on Twitter in line with its Terms and Conditions; for the nature of the information posted and the platform used we could assume that the participants expected the virtual space used to be public; for concerns related to users' anonymity, we did not use personal identifiers, and when we use some quotes as examples we are slightly changing the wording to make the tweets unrecognizable (Social Data Lab, 2019). Our research plans were approved by the Research Ethics Committee of the University of Southampton (ERGO2 no 62176).

Data analysis

The dataset of 3,127 tweets belonging to 2,021 accounts was thematically analysed with the support of Netlytic software (Gruzd, 2016), which is an open source, community-supported text and social networks analyser. To guide our work, we first automatically extracted the 100 words most commonly present in the tweets using Netlytic's text analyser which summarises and visualises social media data (see Figure 1). Please note that Figure 1 also shows the prevalence of the words (i.e., the small numbers near each word).

Figure 1 - Most used words in the dataset

[Figure 1 about here]

Netlytic allows researchers to click on each of the words extracted through Netlytic to access the tweet in which the word is embedded, enabling coding of the context in which the word was used for subsequent thematic analysis. This computational support rendered our qualitative analysis of the large dataset manageable. To enhance the objectivity and accuracy of analysis, all three researchers participated in performing Keyword in Context (KWIC) analysis – an established approach initially used in the machine indexing literature (Luhn, 1960), and more recently adapted by social scientists to analyse large corpora of social media datasets (e.g., Wiedemann, 2016; Miller, 2021; Ugwudike and Sánchez-Benitez, 2022). The analysis involved, (1) deriving codes from the data by analysing the 100 extracted words in the context (tweets) from which they emerged, (2) agreeing the codes, and (3) assembling the codes into themes depicting primary and secondary harms (thematic analysis). The codes agreed were: app attribution; outsourcing and profiteering; NHS links; nepotism; distrust; surveillance;

blame towards named politicians; expensive incompetence; app as delayed and defective; the affiliate is a private company; privatisation: data breaches and GDPR; neoliberalism and individual responsibility; fraud risk; deliberate malfunction; risks of non-compliance; freedom of movement; and covid denialism.

As we demonstrate below, we identified several themes, associated with, (1) forms of expressions of resistance to the perceived primary harms of the government's introduction of the app, and (2) the potential secondary harms that emerged during said resistance (as summarised in Table 1 below).

The primary harms were: false advertisement; institutionalised nepotism; incompetence; and privacy violations; as well as enhanced surveillance and law enforcement. The secondary harms produced by the resisters were: malinformation (when genuine information is shared, but it can potentially cause harm, for instance because it is shared in incomplete or potentially misleading ways); mal/disinformation (false information is shared, willingly or unwillingly, in a way that can cause harm); and conspiratorial thinking; all of which are dimensions of information pollution.[4] They can also all be considered socially harmful since, as already noted, they are forms of information pollution that can foment harmful forms of resistance. On this point, rather than posing genuine concerns, or providing evidence-based insights, they typically propagate false, inaccurate or otherwise potentially misleading information. For instance, they divulge non-science-based information, rumours, and even foster the rejection of potentially lifesaving health initiatives (e.g., Chadwick et al., 2018; Lavorgna and Di Ronco, 2019; Lavorgna, 2021b).

Before we discuss the themes in more detail, it is worth noting that the discourses we classified as instances of information pollution were depicted as such purely because they were produced by resisters in a way that can create or potentiate harms by discouraging others from adopting a potentially useful health initiative. The resisters offered no evidence (or only produced

[4] While we did not perform any quantitative analysis on our dataset as it was unnecessary for the scope of our study, it should be noted that, as we focused on the 100 most found KWIC, all the themes derived from our codes can be considered dominant in our full dataset, as it is unlikely that lone views were captured by using this approach. Nonetheless, it would be interesting for future studies on the potential secondary social harms of certain forms of digital resistance, to approach the full dataset with computational tools than can enable them to further unpack the prevalence and distribution of the views identified.

‘evidence’ dismissed by the scientific community) to substantiate their discourses. We make no claims as to the viability, efficiency, or utility of the app. Further, digital forms of resistance and the use of ONSs to share genuine information are certainly to be welcomed. It is what the app symbolises that matters here: it is a public health initiative which, if efficiently developed and adopted in a stage of the pandemic where scientific evidence on the virus was under development and the international scientific community was consistently advocating for caution, contact tracing and physical distancing, could have saved lives. Information pollution in relation to the app, therefore, can be considered harmful, not least because it has the potential to provoke public rejection of a public health intervention in a time of crisis. Regarding the COVID pandemic specifically, studies found evidence of information pollution which could have fuelled resistance to the use of masks, physical distancing and vaccinations – see e.g. Jamison et al., 2020; Lavorgna, 2021b; Sattar and Arifuzzaman, 2021; Lavorgna et al., 2022). Below we present the themes that emerged from our analysis.

Table 1 - Themes depicting primary and secondary harms (selected samples)
[Table 1 around here]

Limitation of the chosen approach

It is important to note that the approach chosen in this study, while useful for shedding light on how ordinary citizens can utilise their agency to resist the perceived harms of powerful actors and at the same time produce the secondary social harm of information pollution, suffers from a number of limitations. Firstly, we have considered a relatively limited sample from a specific social media platform, and one language/geographical area only. As such, it is important to stress to the reader that, from an empirical perspective, this study should be considered exploratory, a starting point for broader debate on the need to uncover new types of potentially socially harmful practices, and to unpack some assumptions within the criminological scholarship on social harms on the concept of ‘power’ in the digital field.

More importantly, it is essential to note – once more – that our approach does not allow us to draw any conclusion on causality and intent: we cannot prove the impact of the tweets observed, and certainly we cannot prove the motivations behind them. We can only note that, collectively, they had a part in shaping (or at least they tried to shape) public discourse in a

potentially harmful way. Digital social movements and resistance are topics that have attracted an increasing number of scholars, who have demonstrated how social media is used by less powerful groups to frame social issues, raise awareness on harms and injustices, and to advocate for social change (e.g., Freelon et al., 2016). While the attention so far has mostly centred on those digital speeches having a positive impact, we cannot ignore that the same dynamics can propagate harms. This, of course, is not about denying the social importance of forms of digital resistance. We contend that it is necessary to critically explore such practices, to distinguish between the positives (to be supported) and the negatives (to be avoided or mitigated).

Results

Resistance focused mainly on the perceived harms surrounding the app. But expressions of resistance produced secondary harms in the form of various dimensions of information pollution, revealing how resistance by less powerful groups can in itself potentially propagate social harms. The primary harms are outlined below along with the potential secondary harms (forms of information pollution) produced by those resisting the primary harms.

False advertisement

For some resisters, their acts of resistance centred on the harms of false advertisement by the government. They rejected the app on the basis that the app developer, politicians, and the Government were engaging in the harmful act of advertising it as an NHS app when it was in fact an outsourced, commercially produced tool (e.g., '*This is not run by the NHS, this is run by Serco!*'). This can amount to malinformation, as from the context it was clear that the scope of the information shared was not to provide a piece of factual information, but to use that piece of information to oppose *per se* said public health initiative. Similarly, resisters also maintained that the app was being falsely advertised as an NHS product so that the NHS could be demonised in the event of app failure, furthering the government's privatisation agenda (e.g., '*The blame game has started'. 'The NHS is up for grabs*'). This discourse gained traction as many other users advanced it and transformed it into a vital part of the general resistance to the app. We classified it as malinformation since the resisters offered no evidence to verify their claims.

In discussing the harms of false advertisement by the government, other resisters propagated further information pollution in the form of mis/disinformation through the claim, for instance, that the app had been released on 'clap for the NHS day' to link the NHS to the app and ensure that the Service would be blamed if app implementation becomes unsuccessful. There was further mis/disinformation via comments to the effect that the app had been deliberately designed to fail, to support a herd immunity agenda (e.g., '*Anyone else think that having the app release day on the same day we clap for the NHS is suspicious? This is another nail in the coffin for the NHS'*). As already noted, studies show that such information pollution in the form of malinformation and mis/disinformation can foment distrust of the government's motives and resistance to public health initiatives, undermining potential uptake by vulnerable individuals and communities who may otherwise benefit from such resources.

Institutionalised nepotism

Acts of resistance also centred on the perceived harms of nepotism and cronyism in the Government. In particular, resisters claimed that the app contract had been allocated to a well-known and allegedly incompetent, close affiliate of a specific government official. The resisters pushing this narrative also claimed that the app had been introduced to bolster the affiliate's profiteering agenda and quest to maximise profit above all other considerations, including human rights and wellbeing. Thus, in resisting the perceived harm of institutionalised nepotism, which would hamper the possibilities to create an efficient and effective product, the resisters were producing unverifiable discourses that could discourage app use. This amounts to information pollution in the form of mis/disinformation. There were also additional instances of information pollution, this time via malinformation expressed through claims that the affiliate (a private company) had a history of data breaches and other serious forms of incompetence when delivering public services (e.g., '*Stop giving contracts to your incompetent friends!*'; '*How can our data be safe with these people in charge?*'; '*This company has an history of fraud!*').

Information pollution in the form of conspiratorial thinking was also evident in the resisters' claims that the app's developers were far right backers of the Government and Vote Leave affiliates (trying to do something '*profitable for their cronies'*; and '*they can use this data to track and trace voters'*). Again, a potential impact of such information pollution is that it could undermine public trust in government policies and discourage compliance with public health initiatives, in this case, the contact tracing app.

Government incompetence and privacy violations

Additional primary harms linked to the app by resisters were the issues of government incompetence and privacy violations. Again, in resisting these harms, the resisters propagated forms of information pollution that were in profusion across the dataset and involved malinformation via descriptions of the app's defects and comments about the failure to test the app before rolling it out, with critical comments directed at the government for wasting taxpayers' money on a defective app (e.g., *'a catastrophic waste of money!*'; '*£12million wasted!*'; '*£103m wasted!*').

Additional examples also centred on the theme of incompetence with some resisters posting comments and evidence depicting the incompetence of the aforementioned affiliate company in charge of app implementation (e.g., '*This incompetence is killing people'*). Relatedly, there were instances of malinformation pertaining to the government's incompetence in the area of data protection (e.g., '*Incompetence yet zero resignations!*'). Indeed, data insecurity and privacy violations were key concerns amongst the resisters, particularly the ones who were able to promote discourses on this issue across conversational networks.

Resisters in this camp expressed their lack of trust in the government's ability to deal with delicate issues of data protection proficiently. Concerns were raised about the potential of the app, which is data-driven, to foment data breaches and violations. For instance, there were speculations regarding compliance with some aspects of the GDPR and concerns that the app required a number of permissions (including Bluetooth accessibility) which could increase the risk of hacking (e.g., '*It breaks GDPR law!*'; '*this is a cybersecurity nightmare, it worked in China but British people cannot agree to this!*'). Several resisters also noted that private access to sensitive data can prompt other cybersecurity concerns or increase exposure to scams. There were claims that people were being asked to enter credit card details for identification and verification which could expose them to cyber harms.

There were also instances of disinformation through references to the possibility that the private company implementing the app will access people's information for nefarious purposes, and that the Government, working with 'big tech' companies such as Google, would misuse users' data by feeding the companies the data they need for their facial recognition software since the current pandemic and face mask restrictions had reduced the availability of such data (e.g.,

'*Face masks have hindered facial recognition, that track and trace app so wanted by the government could harvest new data'*). Some also claimed that the government would release such data to other countries such as Russia for surveillance policies and practices (e.g., '*the app will release all our data to Russia'*). More generally, the app was portrayed as a data harvesting app for malicious agents and motives, denying its use as a public health tool. Further, the app was depicted by some resisters as a feature of neoliberal responsibilisation. In this respect, some resisters maintained that the Government, incapable of dealing with the pandemic effectively and seeking to divest itself of its responsibility for ensuring adequate public health protection and provision, was (through the app) deliberately transferring responsibility for dealing with the virus to the general public (e.g., '*They should take responsibility for its failures, but they shamefully use political covers'*).

Taken together, these narratives, which were produced or endorsed (e.g., via retweets) by users resisting the app and able to widely promote their resistance across conversational networks, can be considered instances of information pollution. Regardless of the concerns, motives, or intent underpinning the narratives in the Twittersphere, together, the forms of malinformation and mis/disinformation encountered were constructed or propagated in a way that intended to sustain a certain narrative, sustaining the rejection of a public health initiative.

Information pollution via perceived harms of enhanced surveillance and law enforcement

The risk of enhanced surveillance and law enforcement was another primary harm attributed to the app. Our analysis also revealed concerns expressed by resisters regarding the capacity of the app to operate as an additional surveillance technology alongside existing tools such as CCTV cameras (e.g., '*a state surveillance app*'; '*this is just another method of mass surveillance*'). These discourses about the primary harm of an initiative introduced by the Government, contained instances of information pollution in the form of conspiratorial thinking. To illustrate this, some users in this category claimed that the app was part of a wider plan to create a private police force for the future (e.g., *'this is one step towards a private police force*'; '*a private contractor that will soon have the authority to tell you to comply or else, a private police to keep us commoners in line'*). Others maintained that data collected from the app would be misused for practices such as covert surveillance or the promotion of a eugenics agenda ('*our data will be in the hands of eugenics enthusiasts*'). Some also stated emphatically that users' personal data could at some point be de-anonymised to enhance surveillance capacities (e.g., '*ministers are discussing how they might de-anonymise users'*). The

Government does however claim that controls reducing or removing the ability to identify app users are integrated in the app and although the app holds personal data (for example, on venues visited), such data are not stored or shared (UK Health Security Agency, 2021). Nevertheless, speculative concerns about enhanced surveillance through the app formed the basis of expressions of resistance and information pollution via disinformation and conspiratorial thinking.

Discussion

In this paper, we built on an interdisciplinary study that brought together criminological and computational expertise to provide an empirical account of how 'less powerful' people digitally exercise their agency to resist the perceived harms of powerful actors, whilst at the same time producing the secondary harm of information pollution. This apparent paradox, we argue, is of fundamental criminological importance, not only because it uncovers new types of potentially socially harmful practices, but also because it demystifies some assumptions within the criminological scholarship on social harms.

In our paper, we focused on online social resistance which has expanded phenomenally in recent years, partly due to the availability of social media affordances which enable anyone with a suitable device and internet access to share their opinions with a global audience in real time and to 'cluster' with likeminded people just as the resisters in our study did in response to the perceived harms of the contact tracing app**.** Studies of social movements have shown that the experience or even the perception of marginality, victimisation, and other forms of disempowerment (or 'powerlessness', in the social harm jargon – see e.g. Pemberton, 2015) can motivate online resistance. Such resistance is typically intended to raise awareness of the harmful implications of unequal power relations across several spheres.

Although we focused on digital resistance by ordinary citizens against perceived harms of government initiatives, other studies of digital resistance have explored resistance against male violence towards women (e.g., Carney, 2016), and against policing (e.g., Bonilla and Rosa, 2015). Studies of movements such as #BLM and #MeToo reveal how historically disempowered groups, such as women and BAME people, utilise social media sites to influence the representation of important social issues, publicise social harms and injustices, and catalyse

social and political change (e.g., Freelon et al., 2016). These studies suggest that, just as we found in our study, ordinary citizens are able to engage in online resistance against the harms of the powerful by developing counter-narratives and influential oppositional discourses (see also Gallagher et al., 2018).

In his critical analysis of digital media and democratic participation, Dahlberg (2015: 861) notes that digital spaces can enable online 'discursive contestation' in the form of 'contestations of power […] strengthening the voice of alternative, marginalized, or otherwise oppressed groups'. This highlights what has been conceptualised as the empowerment model of online knowledge production (Ugwudike and Fleming, 2021). It can be bolstered by social media algorithms which amplify content that appear to be generating significant interest. In the case of Twitter, for example, content which generates interest in the form of likes, replies and retweets are more likely to attract algorithmic amplification via further dissemination to wider audiences (Koumchatzky and Andryeyev, 2017). This is an empowering feature that equips users with digital capital in the form of the power to produce influential online discourses, including those propagating information pollution.

Digital capital is a relatively new form of capital that has gained ascendancy with the advent of the digital age and the worldwide explosion of web 2.0 which enables anyone to freely create and share information to a global audience in real time, engendering new power relations. From a sociological perspective, digital capital has been defined as the skills, competencies and other resources which confer the power and ability to exploit the benefits of technologies such as ONSs (Ragnedda and Ruiu, 2020; van Dijk, 2005). One such benefit is the ability to control the narrative or prevailing discourse about important issues including public health initiatives.

Existing social harm perspectives emphasise other forms of capital, for example, material or financial capital (see Pemberton, 2015), and demonstrate quite persuasively how ownership or access to such capital explains power asymmetries in society. These perspectives show that the individuals, states, and corporations equipped with such capital able to exercise social, economic, and political power, and enjoy a level of impunity when their actions, produce social harms. Less powerful groups lacking such capital are often the ones exposed to social harms, particularly the harms associated with capitalist production and accumulation, harms such as environmental pollution and Health and Safety violations (Pemberton, 2015; Canning and Tombs, 2021).

Unequal access to digital capital also contributes to power asymmetries, with those lacking such capital also tending to be more exposed to digital harms and marginalisation (van Dijk, 2005). Although a key difference between digital capital and material capital is that the former focuses more on the new types of power imbalance that are emerging with the advent of new technologies such as ONSs. Unlike material capital, the power inequalities associated with digital capital are not necessarily linked to differential abilities to exercise social, economic, and political power. Instead, the inequalities stem largely from the ability of the powerful (in this context, those equipped with digital capital) to use digital products in ways that can, *inter alia*, shape knowledge of key social issues, and in some cases, produce – alongside positive effects –harms such as information pollution. That said, those equipped with this form of capital may be of low socioeconomic status and affected by powerlessness in the offline world. Yet, they may retain the capacity to exercise power through their digital capital, hereby empowering them to control knowledge and information production both online and offline.

We have shown that in the instances of resistance observed in the data, some resisters who were ordinary citizens were able to promote discourses that contained forms of information pollution across conversational networks, specifically information pollution disputing the origins and proposed agenda of the contact tracing app. In other words, speeches of resistance, to be welcomed when opposing injustice and corruption, also became the means through which potentially misleading information travelled. In addition, it is worth considering the role of social media algorithms in conferring digital capital on some users by amplifying their digital content, particularly where such content appears to be generating great interest (Koumchatzky and Andryeyev, 2017).

More generally, the emergence of digital capital as a resource that confers the power to produce influential, but potentially harmful, discourses challenges the predominant (in social harm studies) distinctions made between the powerful and the powerless via the lens of neoliberal or other forms of capital. Historically marginalised groups are now able to access digital capital and the power it confers via social media affordances and other digital facilities. In this context, the paper poses a fundamental question for zemiology and, more generally, critical criminologists in the digital era: should we redefine the meaning of capital, or rather, consider different forms of capital, and why?

Indeed, the empowerment model of social media usage which empowers ordinary citizens, such as the resisters in our study, to produce influential discourses, could be displacing the locus of power from its traditional sites, notably amongst the more powerful (e.g., large corporations and nation states). This could entail the extending of power to new areas, including amongst ordinary citizens equipped with digital capital. As we have shown in our study of public resistance to the contact tracing app, in the current digital age, the power to construct potentially harmful but influential information and forms of knowledge no longer resides solely in the powerful elite, disciplines and institutions.

Therefore, although it has been argued that a progressive social harm perspective must be cognizant of the intersections of access to traditional forms of capital, power, and the capacity to produce social harm (Garside, 2013), this must be reconsidered in the digital age. This is because, focusing on how relations of production central to capitalism generate social harm ignores other harms perpetrated by those outside the capitalist establishment. Examples include those with access to social media and other technological affordances that can confer digital capital. The resisters we studied pose an example.

Digital capital can manifest itself as the power to construct dominant or influential discourses (e.g., van Dijk, 2005), and we demonstrate in this paper that this is a type of capital that brings with it novel harms. Much like neoliberal capital, it is a highly unregulated form of capital with freedom of speech often cited to legitimise it. Owners of digital capital may lack the political and economic power associated with forms of capitalism such as neoliberal capitalism, but just as we found in our study, they have the power (acquired for example through social media affordances) to pollute information. They can even monetise such practice whilst causing social harms. Therefore, their motives and justifications must be critiqued and debunked with relevant moderations and regulations established. This is particularly necessary in the context of potentially life-saving public health interventions. Our study found evidence of resisters using their digital capital to propagate harmful information pollution whilst discussing the contact tracing app. Other studies have shown that people with digital capital (e.g., social media influencers) can produce online health misinformation that can persuade and influence others (e.g., Chadwick et al., 2018; Lavorgna and Myles, 2021; Lavorgna, 2021b). This can include historically disempowered groups who, with their digital capital bolstered social media affordances, construct and disseminate discourses designed to resist perceived harms of the powerful and challenge epistemic domination.

Paradoxically, studies have found that although social media platforms can also be disempowering when they amplify the discourses of the more powerful, rendering others less visible (e.g., Ugwudike and Fleming, 2021), the platforms can also heighten the visibility of the historically marginalised, including those resisting perceived harms of the more powerful in society. Therefore, the distribution of power in online spaces is dynamic. Platform algorithms play a part in this through the amplification or suppression of posts (e.g., Keller, 2021). For example, as already noted, posts that appear to generate considerable attention tend to be subject to algorithmic amplification, conferring digital capital on the authors by allowing them to dominate digital discourses. Algorithmic amplification processes can allow individuals and groups that traditionally lack public visibility and voice offline, to produce influential discourses in online spaces. This new voice might not be sufficient to have a seat in the metaphorical table, but it is enough to confer the power to exercise one's agency and produce widely disseminated information.

Conclusions

Our primary focus in this paper has been to show that in the process of exercising their agency through online resistance in the form of producing potentially harmful information, resisters can produce secondary social harm, such as digitally-enabled information pollution, as has been the focus of this study. We have used the example of resistance to public health interventions, specifically the contact tracing app introduced in the UK in the wake of the COVID-19 pandemic. The harms perpetrated by those resisting such intervention have mostly been ignored in criminological literature. In the context of the COVID pandemic, apart from a few exceptions (see, for instance, Lavorgna & Myles 2021; Lavorgna, 2021b; Lavorgna et al., 2021), critical criminologists and zemiologists studying social harms have mostly focused on the governmental failures in responding effectively to the crisis. Key issues addressed include poor healthcare services, discriminatory responses to the pandemic (Vegh Weis and Magnin, 2021), and increased government support for populist policing (Scalia, 2021). We acknowledge that these are fundamental issues deserving disciplinary attention. We therefore recognise why criminological studies of social harms have traditionally focused on the often minimised or dismissed harms of the powerful, including state harms.

Nevertheless, we contend that there is a risk of overlooking the agency of less powerful individuals and groups affected by the harms of the powerful. In our study, we addressed this problem by focusing on those typically categorised as the 'less powerful' (ordinary citizens). We explored their active efforts to resist real or perceived harms pertaining to the COVID-19 digital tracing app. We have in the process, shown how unrecognised or minimised 'secondary' harms can be produced by individuals resisting the harms of 'the usual (powerful) suspects', and we have focused on the harm of information pollution. Such harm and its embedded mechanisms, such as mis/disinformation, misinformation and conspiratorial thinking, can produce socially harmful outcomes even if they are perpetrated by the 'less powerful'. It is also worth noting that 'less powerful' is not always synonymous with 'victimised'. This is particularly the case where the position of some individuals typically considered less powerful becomes ambiguous. An example is when their access to social media affordances equips them with the digital capital to support, and proactively further, potentially dangerous information pollution.

Acknowledgements

We would like to thank the Reviewers for their valuable feedback.

References

Agarwal, P., Sastry, N., and Wood, E. (2019) 'Tweeting MPs: Digital Engagement between Citizens and Members of Parliament in the UK', *Proceedings of the International AAAI Conference on Web and Social Media*, vol 13, no 1, pp 26-37.

Allington, D., Duffy, B., Wessely, S., Dhavan, N., and Rubin, J. (2021) 'Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency', Psychological Medicine, vol 51, no 10, pp 1763-1769.

Alothali, E., Zaki, N., Mohamed, E.A. and Alashwal, H. (2018) Detecting social bots on twitter: a literature review. In 2018 International conference on innovations in information technology (IIT), IEEE.

Bangerter, A., Wagner-Egger, P. and Delouvee, S. (2020) 'How conspiracy theories spread', in Butter, M. and Knight, P. (eds.), *Routledge Handbook of Conspiracy Theories*, London: Routledge.

Biddlestone, M., Green, R. and Douglas, K.M. (2020) 'Cultural orientation, power, belief in

conspiracy theories, and intentions to reduce the spread of COVID-19', *British Journal of Social Psychology*, vol 59, pp 663-673.

Bonilla, Y. and Rosa, J. (2015), #Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist*, 42(1), 4–17.

Canning, V. and Tombs, S. (2021) *From Social Harm to Zemiology: A Critical Introductio*n, London: Routledge.

Carney, N. (2016). All Lives Matter, but so Does Race: Black Lives Matter and the Evolving Role of Social Media. *Humanity & Society*, 40(2), 180-199.

Carr, L. (2020), 'Web Data Research Assistant' (Version 3.5.11)', https://chrome.google.com/webstore/detail/web-data-research-assista/kcdbekkmigohaijilebpaeoopcgjbbdm (accessed October 2020)

Chadwick, A., Vaccari, C., O'Loughlin, B. (2018). Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. *New Media & Society*, 20(11), 4255–4274.

Coday, D. (2006) 'An introduction to the philosophical debate about conspiracy theories', in Coday, D. (ed.), *Conspiracy theories. The philosophical debate*, London: Routledge.

Cosi, G. (1984) *Saggio sulla disobbedienza civile. Storia e critica del dissenso in democrazia*. Milan. Giuffrè.

Dahlberg, L. (2011), Re-Constructing Digital Democracy: An Outline of Four 'Positions', *New Media and Society*, 13(6): 855–872.

Douglas, K.M., Cichocka, A. and Sutton, R.M. (2020) 'Motivations, emotions and belief in conspiracy theories', in Butter, M. and Knight, P. (eds.), *Routledge Handbook of Conspiracy Theories*, London: Routledge.

Durkheim, É (1893/1984) *The division of labor in society*. New York: The Free Press.

Durkheim, É. (1900/1969) 'Two laws of penal evolution', *University of Cincinnati Law Review*, 38, 32–60.

Ferrari, V. (2004) *Diritto e società. Elementi di sociologia del diritto*. Rome: Editori Laterza.

Ferrell, J. (2019) 'In Defense of Resistance', Critical Criminology, (2019). https://doi.org/10.1007/s10612-019-09456-6.

Ferri, E. (1921) 'Reform of Penal Law in Italy'. *Law Quarterly Review*, 37, 63-85.

Freelon, D., McIlwain, C., Clark, M. (2016) Quantifying the Power and Consequences of Social Media Protest. *New Media & Society*. https://doi.org/10.1177/1461444816676646

Gallagher, R., J., Reagan, A.J., Danforth, C.M., and Dodds, P.S. (2018) Divergent discourse

between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. *PLoS ONE* vol 13, no 4).

Garside, R. (2013). Abordar el daño social: mejor regulación o transformación social (Addressing social harm: better regulation versus social transformation) *Revista Crítica Penal Y Poder*, 5, 225–246.

Gordon, F., McGovern, A., Thompson, C., and Wood, M.A. (2022) 'Beyond cybercrime: New perspectives on crime, harm and digital technologies', *International Journal for Crime, Justice and Social Democracy*, 11(1).

Greenfield, V.A., and Paoli L. (2013) 'A framework to assess the harms of crimes', *The British Journal of Criminology* vol 53, no 5, pp 864-685.

Gruzd, A. (2016). Netlytic: Software for Automated Text and Social Network Analysis. Available at http://Netlytic.org.

Hill, M.L. (2018) '"Thank you, Black Twitter": State violence, digital counterpublics, and pedagogies of resistance', *Urban Education*, 53(2), 286-302.

Hillyard, P., & Tombs, S. (2004) Beyond Criminology? In Hillyard, P., Pantazis, C., Tombs, S., & Gordon, D. (Eds.), *Beyond Criminology: Taking Harm Seriously* (pp. 10–29). London: Pluto Press.

Hillyard, P. and Tombs, S. (2017) Social Harm and Zemiology, in Liebling, A., Maruna, S. and McAra, L. (eds) *Oxford Handbook of Criminology* (6th edn). Oxford: Oxford University Press).

Hopmann, D.N. and & Shehata, A. (2011) 'The Contingencies Of Ordinary Citizen Appearances In Political Television News', *Journalism Practice*, vol 5, no 6, pp 657-671.

Imhoff, R. and Lamberty, P. (2020) 'Conspiracy beliefs as psycho-political reactions to perceived power', in Butter, M. and Knight, P. (eds.) *Routledge Handbook of Conspiracy Theories*, London: Routledge.

Jamison, A. M., Broniatowski, D. A., Dredze, M., Sangraula, A., Smith, M. C. and Quinn, S. C. (2020) 'Not just conspiracy theories: Vaccine opponents and proponents add to the COVID-19 'infodemic'on Twitter', *Harvard Kennedy School Misinformation Review*, 1.

Keller, D. (2021). Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard. J. FREE SPEECH L., 1, 227-268.

Khan, S.R. (2012) 'The Sociology of Elites', *Annual Review of Sociology*, vol 38, pp 361-377.

Krippendorff, K. and Halabi, N. (eds) (2020) *Discourses in Action: What Language Enables Us to Do*. London: Routledge.

Koumchatzky, N. and Andryeyev, A. (2017) Using deep learning at scale in Twitter's timelines. Available at: https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html (accessed September 2020).

Lasslett, K. (2010) 'Crime or social harm? A dialectical perspective', *Crime Law Soc Change*, 54, 1–19.

Lavorgna, A., and Di Ronco, A. (eds.) (2019) *Medical Misinformation and Social Harm in Non-Science Based Health Practices: A Multidisciplinary Perspective.* London: Routledge.

Lavorgna, A. (2021a) Looking at Crime and Deviancy in Cyberspace Through the Social Harm Lens, in Davies, P., Leighton, P. S. and Wyatt, T. (eds) *The Palgrave Handbook of Social Harm*. Basingstoke: Palgrave Macmillan.

Lavorgna, A. (2021b) *Information pollution as social harm: Investigating the digital drift of medical misinformation in a time of crisis*. Bingley: Emerald.

Lavorgna, A. and Myles, H. (2021) 'Science denial and medical misinformation in pandemic times: A psycho-criminological analysis', *European Journal of Criminology*, doi:10.1177/1477370820988832.

Lavorgna, A., Ugwudike, P., Carr, L., Benitez, Y.S. and Rekha, G.S. (2021a) 'To app or not to app?: Understanding public resistance to COVID-19 digital contact tracing and its criminological relevance', *Law, Technology and Humans*, vol 3, no 2, pp 28-45.

Lavorgna, A., Ugwudike, P., Sanchez Benitez, Y. and Carr, L. (2021b) *Understanding public resistance in using COVID-19 digital contact tracing app: an interdisciplinary analysis*. Project report & policy brief, University of Southampton. Available at: https://www.southampton.ac.uk/research/projects.

Lavorgna, A., Carr, L. and Kingdon, A. (2022) 'To wear or not to wear? Unpacking the #NoMask discourses and conversations on Twitter'. Under Review.

Leighton, P. (2021) The Harms of Industrial Food Production: How Modern Agriculture, Livestock Rearing and Food Processing Contribute to Disease, Environmental Degradation and Worker Exploitation, in Davies, P., Leighton, P. S. and Wyatt, T. (eds) *The Palgrave Handbook of Social Harm*. Basingstoke: Palgrave Macmillan.

Lloyd, A. (2021) Work-Based Harm, in Davies, P., Leighton, P. S. and Wyatt, T. (eds) *The Palgrave Handbook of Social Harm*. Basingstoke: Palgrave Macmillan.

Luhn, H.P. (1960) 'Key word- in- context index for technical literature (kwic index)', *American documentation*, vol 11, no 4, pp 288-295.

Miller, D.T. (2021) 'Characterizing QAnon: Analysis of YouTube comments presents new conclusions about a popular conservative conspiracy', *First Monday*.

Nielsen, J. (2003) 'IM, Not IP (Information Pollution)', *ACM Queue*, vol 1, no 8, pp 75-76.

Ofori-Parku, S.S. and Moscato, D. (2018) 'Hashtag Activism as a Form of Political Action: A Qualitative Analysis of the #BringBackOurGirls Campaign in Nigerian, UK, and U.S. Press', *International Journal of Communication*, vol 12, no 23.

ONS (2020) *Why have Black and South Asian people been hit hardest by COVID-19?* https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/whyhaveblackandsouthasianpeoplebeenhithardestbycovid19/2020-12-14

Pemberton, S. (2015), *Harmful societies. Understanding social harms*. Bristol: Policy Press.

Peter, C. and & Zerback, T. (2020) 'Ordinary Citizens in the News: A Conceptual Framework', *Journalism Studies*, vol 21, no 8, pp 1003-1016.

Presser, L. (2013), Why We Harm, New Brunswick, NJ: Rutgers University Press.

Pummerer, L., Böhm, R., Lilleholt, L., Winter, K., Zettler, I. and Sassenberg, K. (2022) 'Conspiracy Theories and Their Societal Effects During the COVID-19 Pandemic', *Social Psychological and Personality Science*, vol 13, no 1, pp 49-59.

Ragnedda M., Ruiu M.L. and Addeo F. (2020) 'Measuring Digital Capital: An empirical investigation', *New Media & Society*, vol 22, no 5, pp 793-816.

Romer, D. and Hall Jamieson, K. (2020) 'Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S.', *Social Science & Medicine*, 263.

Rothe, D., and Kauzlarich, D. (2016) *Crimes of the powerful*. London: Routledge.

Royal Society for Public Health (2020) *Public attitudes to a Covid-19 vaccine, and their variations across ethnic and socioeconomic groups.* https://www.rsph.org.uk/static/0409f277-9ec5-4bde-81e3c07cff95d746/COVIDvaccinePaper4.pdf).

Sattar, N.S. and Arifuzzaman, S. (2021) 'COVID-19 vaccination awareness and aftermath: Public sentiment analysis on Twitter data and vaccinated population prediction in the USA', *Applied Sciences*, vol 11, n 13, pp 6128.

Schwendinger, H. and Schwendinger, J. (1977) ,Social class and the definition of crime', *Crime and Social Justice*, 7, 4–13.

Scalia V. (2021) "Stay home you murderer!": populist policing of COVID-19 in

Italy', *International Journal of Police Science & Management*, vol 23, no 3, pp 242-252.

Smith, G.D.J. and O'Malley, P. (2017) 'Driving Politics: Data-driven Governance and Resistance', *The British Journal of Criminology*, vol 57, no 2, pp 275-298.

Social Data Lab. (2019) *Lab online guide to social media research ethics.* Available at: http://socialdatalab.net/ethics-resources.

Sutherland, E.H. (1983) *White collar crime: The uncut version*. New Haven: Yale University Press.

Tift, L., L., and Sullivan, D. C. (2001) A Needs-Based, Social Harms Definition of Crime, in Henry, S. and Lanier, M. (eds) *What is Crime? Controversies over the Nature of Crime and What to do About It*. Lanham: Rowman and Littlefield Publishers, Inc.

Tombs, S. (2014), 'Hard Evidence: are work-related deaths in decline?', The Conversation, 29 October, https://theconversation.com/hard-evidence-are-work-related-deaths-in-decline-33553)

Ugwudike, P. (2015) *An introduction to critical criminology*. Bristol: Policy Press.

Ugwudike, P. (2020) 'Digital prediction technologies in the justice system: The implications of a "race-neutral" agenda', *Theoretical Criminology*, vol 24, no 3, pp 482-501.

Ugwudike, P. (2021) Data-Driven Technologies in Justice Systems: Intersections of Power, Data Configurations, and Knowledge Production, in Lavorgna, A. and Holt, T. (eds) *Researching Cybercrimes*, Cham: Palgrave Macmillan, Cham.

Ugwudike, P., and Fleming, J. (2021) 'Artificial intelligence, digital capital, and epistemic domination on Twitter: A study of families affected by imprisonment', *Punishment & Society*, 14624745211014391.

Ugwudike, P. and Sánchez-Benitez, Y. (2022) 'Critical Social Media Analysis: Problematising Online Policy Representations of the Impact of Imprisonment on Families', *International Journal of Offender Therapy and Comparative Criminology*, doi:10.1177/0306624X221086559.

UK Health Security Agency (2021) NHS COVID-19 app: anonymisation, definitions and user data journeys. Available at: https://www.gov.uk/government/publications/nhs-covid-19-app-privacy-information/anonymisation-definitions-and-user-data-journeys

Van Dijk J (2005) *The deepening divide: Inequality in the information society. Thousand Oaks*, CA: Sage.

Vegh Weis, V., and Magnin, B. (2021) 'Essential Crimes? Essential Punishments?

Rethinking Essentiality in the Midst of the COVID-19 Pandemic', *Critical Criminology*, vol 29, pp 273-288.

Wardle, C. and Derakhshan, H. (2017) *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making.* Strasbourg: Council of Europe.

Wardle, C (2018) *Information disorder: The essential glossary*. Harvard, MA: Shorenstein Center on Media, Politics, and Public Policy, Harvard Kennedy School.

White, R. (2021) Global Harms and the Natural Environment, in Davies, P., Leighton, P. S. and Wyatt, T. (eds) *The Palgrave Handbook of Social Harm*. Basingstoke: Palgrave Macmillan.

Wiedemann, G. (2016) *Text mining for qualitative data analysis in the social sciences (Vol. 1).* Wiesbaden: Springer.

Williams, S. (2015) 'Digital Defense: Black Feminists Resist Violence with Hashtag Activism', *Feminist Media Studies*, vol 15, no 2, pp 341-344.

Wood, M. A. (2021). Mapping technology-harm relations: From ambient harms to zemiosis. *Crime, Media, Culture*. https://doi.org/10.1177/17416590211037384

World Health Organization (2020) Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. Joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC. https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation

Xiong, Y., Cho, M., and Boatwright, B. (2019) 'Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement', *Public Relations Review*, vol 45, no 1, pp 10-23.

Yang, G. (2016) 'Narrative Agency in Hashtag Activism: The Case of #BlackLivesMatter', *Media and Communication*, vol 4, no 4, pp 13-17.

Yar, M. (2012) 'Critical criminology, critical theory and social harm', in Hall, S. and Winlow, S. (eds) *New Directions in Criminological Theory*. Taylor & Francis.

Ziccardi, G. (2012) Resistance, liberation technology and human rights in the digital age (Vol. 7). Cham: Springe.

Zimmer, M., and Kinder-Kurlanda, K. (eds.) (2017). *Research ethics for the social age. New challenges, cases, and contexts*. New York: Peter Lang.

Tables & Figures

Figure 1 - Most used words in the dataset

Table 1 - Themes depicting primary and secondary harms (selected samples)

Primary Harms *(Perceived harms of the powerful)*	**Secondary Harms** *(Types of information pollution by those resisting primary harms)*
False advertisement	Malinformation • *App is privatised/ commercial app. but advertised as an NHS tool* Mis/Disinformation • *The app has been deliberately designed to fail, to support a herd immunity agenda.*
Institutionalised nepotism	Mis/Disinformation • *App has been introduced to bolster the profiteering agenda of a close affiliate of the current government* Conspiratorial Thinking • *Developers of the app are far right backers of the Government and Vote Leave affiliates*
Government incompetence and privacy violations	Malinformation • *App will foment data breaches and violations.* Mis/Disinformation • *App grants private companies access to confidential data for nefarious purposes.*

Enhanced surveillance and law enforcement	Conspiratorial Thinking • *App data will be misused for practices such as covert surveillance or a eugenics agenda.*