UNIVERSITY OF Southampton

Understanding public resistance to COVID-19 digital contact tracing apps

A sociotechnical analysis

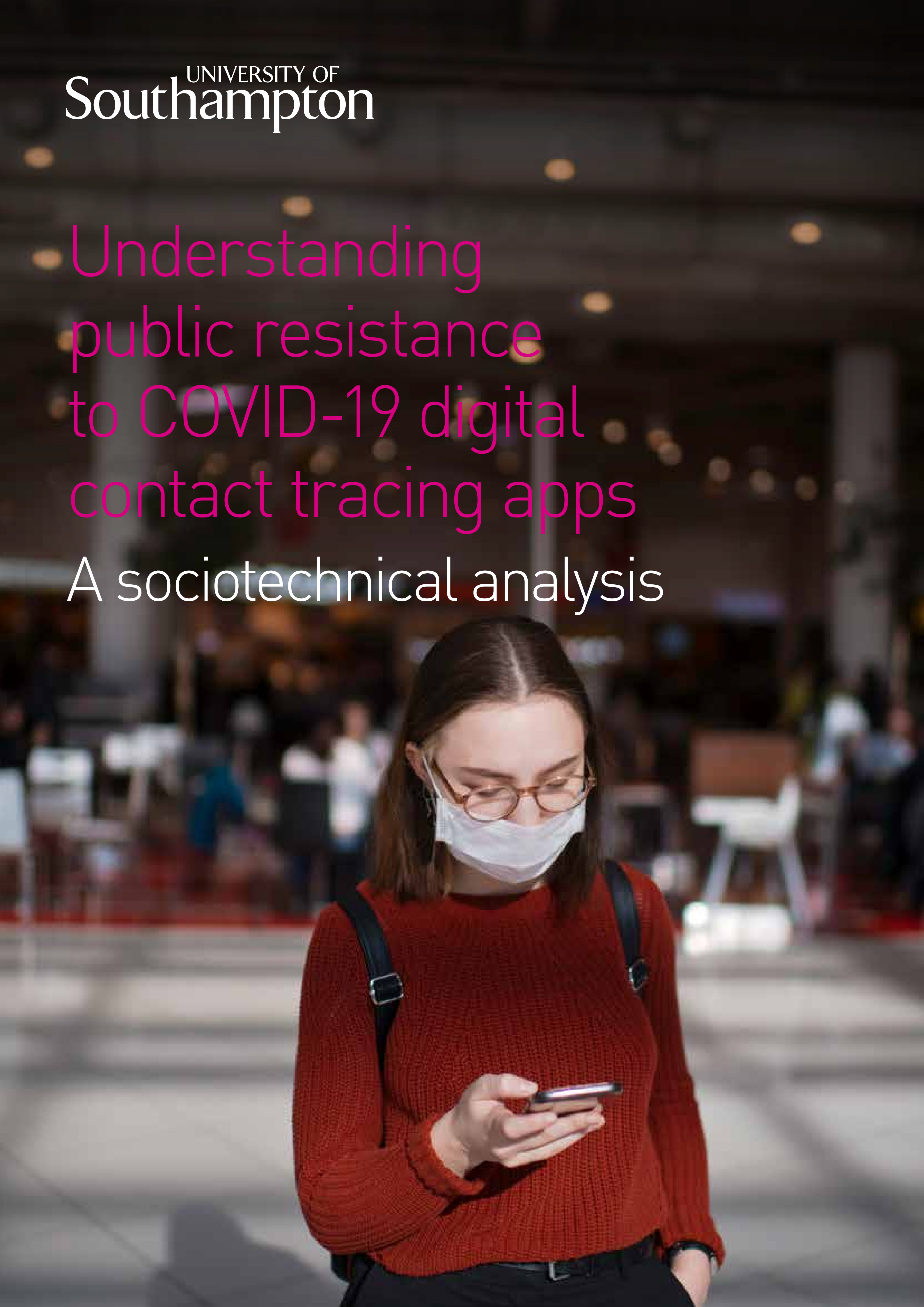


TABLE OF CONTENTS

Acknowledgments

We thank Ms Sasia Gopala Rekha for her support with collecting literature review material and Ms Areej Alzaidi for work on data collection and analysis.

Funding

This research has been funded by the WSI Research Collaboration Stimulus Fund 2020, University of Southampton.

BACKGROUND

Contact tracing apps in the context of the pandemic

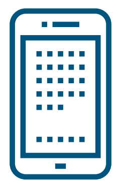
In early 2020, the world found itself facing a new challenge, with the outbreak of a novel coronavirus disease – COVID-19, which was first identified in December 2019 in the Hubei province of China, spreading across countries to the point that the outbreak was recognized as a pandemic by the World Health Organization (WHO) on 11 March 2020. As we all know, the impact on public health of the pandemic has been – and still is – very serious. Alongside medical responses, governments worldwide have applied a number of measures such as lockdowns, use of face masks, widespread physical distancing measures, and active case finding (testing and isolation, contact tracing, and quarantine) to attempt to curb transmission.

Contact tracing apps were developed and released in several countries, including the United Kingdom, as a measure to combat COVID-19, speeding up the tracing of contacts of people found to be infected (Ada Lovelace Institute, 2020). The NHS COVID-19 contact tracing app was launched in England and Wales at the end of September 2020 to help control coronavirus transmission. At the time of writing, the app has been downloaded on more than 21 million phones, with about 16.5 million regular users which is roughly 28 percent of the population (Lewis, 2021) and although numbers in the United Kingdom have been good, they are not impressive, according to researchers (Wymant et al., 2021).

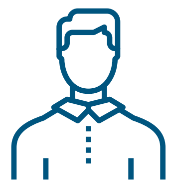
Studies show mounting evidence that apps can help prevent infections and are a valuable public health tool (e.g., Lewis, 2021). For example, researchers evaluated the NHS COVID-19 app from its launch in September 2020 to the end of December 2020. The evaluation found that the app was used regularly by approximately 16.5 million users (28 percent of the total population), and sent approximately 1.7 million exposure notifications (Wymant et al., 2021). Similarly, a pilot study in Spain found that the COVID-19 tracing app notified roughly twice the number of people exposed, compared with manual contact tracing (Rodríguez et al., 2021).

Despite mounting evidence showing the usefulness of digital contact tracing apps, researchers have also identified barriers to an app's effectiveness, such as how well the app is integrated into the local health-care system (Marjanovic et al., 2020) and/or how they measure exposure risk (Masel et al., 2021).

Cautions against expanding an app's functions beyond what the public will accept, is also central to researchers' concerns (Lewis, 2021). Privacy, for instance, has been at the centre of COVID-19 contact tracing app concerns with unprecedented levels of surveillance, data exploitation and misinformation being tested across the world since the beginning of the pandemic (Privacy International, 2021).



21
million downloads



16.5
million users

1.7
million notifications

Challenges and limitations to using contact tracing app technologies

The future of public health is likely to become increasingly digital, and the COVID-19 pandemic for the first time 'tested', on a global laboratory, the ability of apps to help control and prevent infections, providing that they have adequate political backing and are properly integrated into public health systems (Lewis, 2021).

However, there are key challenges and issues that go beyond political decisions and integration. These challenges influence the willingness to use these contact tracing apps, and include ethical issues such as the potential abuse of user privacy rights, lack of trust in the government and public health authorities, security vulnerabilities, user behaviour and participation, and technical constraints (Akinbi et al., 2021).

In terms of technical constraints, these apps are generally based on practical hardware technologies (e.g., Bluetooth low energy, possibly GPS data), so basically anyone with a smartphone can use and implement them. In practice, however, these types of apps lack sufficient real-life testing, which is problematic as their effectiveness, regardless of the technology used, depends, *inter alia*, on socio-behavioural factors such as public confidence and trust in the protection of privacy (Sweeney, 2020; Lavorgna et al., 2021a).

Research carried out over the course of the pandemic, for instance, has shown that privacy and data security are core concerns of the general population in many countries (Farronato et al., 2020; Fitriani, 2020), with further concerns around transparency about how the app works, and how data are collected, protected, stored and shared (Alsdurf et al., 2020), and the potential misuse of the app technology for surveillance purposes (Farries, 2020; Garret et al., 2020). Abeler and colleagues' study during the pandemic (2020) found out that there was wide support for app-based contact tracing in the United Kingdom (with about three-quarters of respondents saying that they would definitely or probably install the app, comparatively higher than in other countries); but respondents who lacked trust in the government were less favourable due to fears of government surveillance and hacking.

Another challenge in the deployment of contact tracing apps has been the lack of alignment of international strategies for the regulation, evaluation and use of digital technologies (Budd et al., 2020). Due to myriad concerns over data privacy, several proposals on achieving digital contact tracing governance are also discussed (Okamoto & Fujita, 2020), alongside comprehensive regulatory systems (Farronato et al., 2020).

As such, trust in technological solutions (Shin, 2020; Shin et al,. 2020) and in the authorities encouraging adoption (including the Government) are seen to be a necessary condition for the implementation of technologies that require public compliance. Public distrust of algorithms (automated-decision making), of their purpose and effectiveness, and of the decisions they make, has been reinforced by growing cases of the harms of data-driven decision technologies (see Angwin & Larson, 2016) as well as growing awareness of ethical issues such as privacy violations, limited explanability, transparency, and accountability (Felzmann et al., 2020; Lavorgna & Ugwudike, 2021).

Broader ethical concerns and harms – digital exclusion and algorithmic injustice

There are other important elements of concern that need to be recognised and discussed. As the pandemic unfolded, questions emerged around whether there was sufficient understanding of contact tracing apps and whether there could be a balance between protecting public health with safeguarding civil rights (Sweeney, 2020).

The debate surrounding contact tracing apps seemed to primarily focus on centralized (anonymized data on a central server) versus decentralized (data distributed on individual devices) approaches, becoming an argument over technical architectures which tends to techno-solutionism of social problems (Milan, 2020) and issues of privacy (as mentioned before). These concerns, however, have partially led to the marginalization of other important issues such as how people who failed to install these apps might be discriminated against, especially those who are already vulnerable or who do not have sufficient agency (Sánchez-Monedero et al., 2020).

As the pressure to deploy automated decision-making systems in the public sector become prevalent, the current data governance regimes and national regulatory practices can be intensifying existing power asymmetries (Kuziemski & Misuraca, 2020) that may harm certain populations by forms of digital exclusion and unfair outcomes.

Concerns around digital exclusion have been emphasised, including in the Ada Lovelace Institute report (2020) which found a 'data divide' of inequalities in access, knowledge and awareness of digital health technologies used in the pandemic, including contact tracing apps. According to the digital inclusion organization Good Things Foundation, almost 2 million British households do not have any internet access while a further 7 million people have used the internet but have no knowledge of how to open an app (Hern, 2020).

Some populations that are particularly vulnerable to the health impacts of COVID-19 (e.g., older adults, people who are homeless, and socioeconomically deprived populations, for instance those with precarious legal status or running from state violence) are also less likely to own a smartphone, potentially amplifying their risks because contact-tracing apps could – for similar reasons – be less likely to reduce transmission within their social circles.

Corresponding equity issues in data-driven approaches to medicine can, for example, arise in electronic health records (EHR) that fail sufficiently to include members of disadvantaged or marginalized groups who are unable to access the health care system (Arpey et al., 2017; Gianfrancesco et al., 2018), or in the sample selection biases that emerge when data availability is limited to well-resourced, digitally mature hospitals that disproportionately serve a particular racial or socioeconomic segment of a population to the exclusion of others.

Critical questions and responses should also concern the power and control different actors hold over these technologies and the data they yield, which may include not only public health authorities, but other governmental agencies (e.g., police, immigration), universities, technology companies (data hosting platforms, software and social media networking platforms), and other players (health insurers, retailers, data brokers) (Pagliari, 2020).

The sociotechnical complexities around implementing contact tracing apps underlines the need for holistic, inclusive and adaptive strategies that take into account the public's context, agency and control over their own data (Pagliari, 2020; Hogan et al., 2021). Studies grounded in the social sciences, for instance, have recently found that resistance mechanisms to using contact tracing apps may arise such as conspiratorial thinking (Lavorgna et al., 2021a; Lavorgna & Myles, 2021) as they offer some compensatory sense of control, agency and power by rejecting official narratives (Douglas et al., 2020; Imhoff & Lamberty, 2020; Lavorgna, 2021).

The role of social media in amplifying misinformation and (algorithm) distrust

Concerns around misinformation (and, similarly, other forms of 'information pollution' such as disinformation and malinformation, see Wardle & Derakhshan, 2017; Lavorgna, 2021) in social media platforms have been central to the power and control these platforms have in the handling of the pandemic (Gonzalez-Padilla & Tortolero-Blanco, 2020; Salvi et al., 2021).

Social Network Sites (SNSs) have been shown to be the fastest and most convenient type of online platforms used to search for and circulate information on specific issues (Lee & Choi, 2018). They are also broadly used by online users to address and confirm specific rumours and facts during uncertain situations. Nowadays it is well demonstrated that SNSs are a powerful medium for the circulation of large volumes of non-supervised journalistic content (Lazer et al., 2018), empowering both (unwilling) misinformation and (malicious) disinformation, and thus provoking the possibility of manipulating the public's perception of the reality through the spread of fake or otherwise misleading news (Ireton et al., 2018). Scholars who have explored how SNSs have been used to improve or reduce trust in scientific expertise (including during the COVID-19 pandemic) have highlighted the capacity for social media to be deployed as mechanisms of misinformation, to undermine both public trust in scientific expertise and accompanying systems such as algorithms (see, among many others, Llewellyn, 2020; van Dijck & Alinejad, 2020).

An example of social media mechanisms that play a role in public misinformation is the infrastructure behind Twitter content management. Twitter relies on a deep learning algorithm that has learned to prioritize content with greater prior engagement (Koumchatzky & Andryeyev, 2017). By combing through Twitter's data, the algorithm has taught itself that Twitter users are more likely to engage with content that has already received many retweets and mentions, compared with content that has fewer. So, if a tweet is retweeted, favourited, or replied to by enough of its early viewers, the timeline algorithm will show it to more users, at which point it will tap into the biases of those users too – prompting even more engagement, and so on. At its worst, this cycle can turn social media into a kind of confirmation bias machine (Tufecki, 2017), one perfectly tailored for the spread of misinformation.

The engineering architecture of social media networking platforms indicates that technological companies have direct power and control over how information is processed, seen, engaged with and spread (e.g., Ugwudike & Fleming, 2021). This raises questions over who controls what and is central to concerns over algorithmic harms that arise from such computational form and consequent distrust in technologies (algorithm distrust). As discussed in Lavorgna (2021), in lack of more profound architectural changes, interventions from online intermediaries, targeting the source, are proving relatively ineffective, and can potentially create serious tensions against individual rights.

Unfortunately, polluted information and, in primis, misinformation phenomena during the management of disease outbreaks, speed up the epidemic process by influencing and fragmenting social response (Kim et al., 2019), and we witnessed this problem during the ongoing pandemic. As we will see, these mechanisms had an effect on the adoption of the contact tracing app.

ADDRESSING THE CHALLENGE - A SOCIOTECHNICAL APPROACH

Our project

Our project "*To app or not to app? Understanding public resistance in using COVID-19 digital contact tracing*", relies on an interdisciplinary approach bringing together socio-criminological and computational expertise. The project aimed to further understand the challenges and concerns using contact tracing apps from the public's point of view by investigating a large Twitter dataset to unravel the social dynamics underpinning people's resistance to the NHS contact tracing app across England and Wales.

Twitter data, as many other comparable social media data, can be seen as qualitative data but on a quantitative scale; novel methodological approaches can be used alongside the traditional tools of social science researchers to make better, more comprehensive sense of such data. As we will see, the interdisciplinary, sociotechnical approach adopted in this study proved useful to ethically study online networks and their discourses at both sufficient breadth and depth.

The understanding of public resistance to using contact tracing apps is inherently sociotechnical. An important part of the application of sociotechnical systems (STSs) is the development of methods, tools and techniques to assess human factors of a technological artifact (Waterson et al., 2016).

Researchers and policymakers recognise that even innovations with a rigorous and proven evidence base often fail to achieve uptake and spread (NHS Digital, 2017; Greenhalgh & Papoutsi, 2019). A recent study of Marjanovic and colleagues (2020) in England highlights that efforts to ensure sustainable innovating healthcare systems require attention to multiple aspects of the system simultaneously, including to behavioural and cultural levers, as much as to the more commonly targeted technical and structural interventions.

The sociotechnical approach of our study contributes to advancing the applications of complex systems thinking to major social transformation challenges (in our case, the ongoing pandemic) by using a combination of computational tools (i.e., data collection software, sentiment analysis, Natural Language Processing methods and network analysis) and criminological analytical expertise that allows for better understanding of social behaviours.

Insights from Criminology

As mentioned in previous sections above, concerns around adopting digital contact tracing technologies tend to focus on the technical architectures around these technologies without much emphasis on users' agency and control over how these technologies operate and what this means to the users. This opens an opportunity to analyse and understand resistance to adoption of contact tracing technologies from the public's point of view. Thus, our study sought to enhance criminological understandings of the factors underpinning resistance to official techs such as digital tracing apps and identify remedial strategies.

From a criminological standpoint, our study of the digital tracing app uncovers new insights that can expand current understandings of resistance to the new data-driven surveillance technologies currently transforming the landscape of decision making across the private and public sector, including the justice system. The problem of resistance to official, policy driven techs is of great relevance to the discipline, particularly to the fast-growing strand of criminology that focuses on the design and adoption of emerging data-driven technologies, some of which include the rapidly proliferating predictive algorithms.

However, the extant criminological literature has to date, mostly focused on coercive surveillance systems such as electronic tags (e.g., Nellis, 2015) and CCTV surveillance systems (e.g., Gannoni et al. 2017), digital prediction technologies in the criminal justice system (e.g., Ugwudike, 2020; Završnik, 2020), and sociotechnical dynamics and barriers to equitable knowledge production (e.g., Ugwudike & Fleming, 2021; Ugwudike, 2021). Our study aimed to further this research thread by investigating a surveillance tech that relies on public acceptance and voluntary adoption for effective deployment, hence filling an important gap in the literature.

Insights from AI Design literature

The study draws on sections of the AI design literature that explore User Experience (UX) of data-driven technologies. Originating initially in industry settings and used by organisations seeking to embed user feedback in tech design, Human Computer Interaction (HCI) and UX studies generate information required for developing responsive user-friendly systems. A clear and broad theme that emerges from Human Computer Interaction and UX studies (Blaynee et al., 2016) is the fact that user endorsement of the functionality, utility, usability, and efficiency of a system is necessary for tech adoption.

Added to this, to encourage uptake even in multi-stakeholder conditions, tech design should be responsive to broader concerns such as the sociocultural contexts of use, including users' entrenched beliefs and interests (Ferreira, 2016). HCI offers helpful insights as the study of how people use technological artifacts, and their design (Stanney et al., 2007); HCI works at the intersection between psychology and the social sciences, on the one hand, and computer science and technology, on the other (Carroll, 1997).

The literature on AI technologies design and ethics offers great insight into the potential challenges and opportunities in technological adoption. Human-centred AI research strategies, for example, emphasize the humanistic and ethical frontiers of AI. Contact tracing apps rely heavily on AI design because its functionality relies on automated recommendations on the basis of data inputs. Resistance to using contact tracing apps represents a failure of adoption. This shows the importance of HCI design for usable AI. A human-centred AI perspective which is inherently interdisciplinary allows us to further unpack the importance of centring the public in the design of technologies for their adoption (Auernhammer, 2020).

Social justice has been a strong focus in recent HCI work (Bellini et al., 2020). On the same line, research from Dombrowski and colleagues (2016) states the importance of so-called Social Justice Oriented Design, suggesting that a key issue is the recognition of 'unjustness' in existing systems. Challenges of technology adoption have been explored, among others, by Verma and Dombrowski's study (2018) into how incorporating data-driven big data techniques into policing technologies design can generate novel problems for both the state citizen.

The HCI community helps in challenging technological design by taking into consideration groups often excluded from design and unfairly targeted and harmed by increasingly complex systems that impose a level of obedience (Bellini et al., 2020, p.3). As such, it is critical that we examine and make explicit the impact of the public's resistance to techs to inform safer, intelligent and just digital and non-digital spaces for all.

Tools from the Web Science

Both Web Science and HCI are intrinsically interdisciplinary fields concerned with the intersection of people and technology. Web Science studies the impact of the Web on society and vice versa, focusing on Web-enabled social practices (Hopper & Dix, 2013). The Web needs to be studied and understood as a phenomenon but also as something to be engineered for future growth and capabilities (Hendler et al., 2008).

An understanding of the sociotechnical interactions enabled by the Web can help in making informed decisions, from government policy, infrastructures and standards, to understanding how social network sites fail to support the richness and dynamism of human relations. At the microscale, the Web is an infrastructure of artificial languages and protocols; a piece of engineering but it is the human beings creating, linking and consuming information that generates the Web's behaviour as emergent properties at the macro scale (Hendler et al., 2008).

Therefore, sociotechnical collaborations are at the very core of Web Science, which has a strong tradition of interdisciplinary research with, among others, both the social and the health sciences. Consider, among others, the work on 'sociodigital' futures (Halford, 2020), the work on tailored e-health interventions to make them work better (e.g., Pope & Turnbull, 2017), and recent work on population level health policies and collaborative data sharing for health and social care transformation (Boniface et al. 2020). These are all interdisciplinary works that rely on Web Science approaches.

Sociotechnical collaborations using Web Science approaches often lead both to computational research and mixed method designs, where computational skills and in-depth, qualitative insights are used in combination trying to offer as much as depth and breadth as possible in investigating a certain phenomenon. In our work, we aimed to progress in this second strand, in line with some recent research carried out by research group (e.g., Alrajebah et al., 2017; Alrajebah et al., 2018; Lavorgna & Carr, 2021; Lavorgna et al. 2021 a,b; Sanchez Benitez, 2021).

Overcoming novel practical and sociotechnical challenges

It has been established that the adoption of technologies relies heavily on the design and social context in which it is deployed. Any social system operating on a technical base (e.g., Twitter, contact tracing apps) are social and technical. If technology design is computing built to hardware and software requirements, then sociotechnical design is computing built to personal and community requirements as well. In sociotechnical systems, the new 'user' of computing is the community (Whitworth, 2009).

Therefore, to limit computing to hardware (engineering) or software (computer science) denies its obvious evolution of social interactions with both hardware and software. HCI began with the personal computing era, so adding people to the computing equation meant that getting technology to work was only half the problem; the other half was getting people to use it (Whitworth & Ahmad, 2013). Just as HCI applies psychology to computing design, current contact tracing computing is another example of a human requirement defining computing.

Similarly, Web Science as an interdisciplinary area can provide us with insights on how the Web developed and keeps developing and how it has affected and is affected by society (Hall & Tiropanis, 2012) and the many social events that are talked about in the Web can offer a view of the public's reactions.

A sociotechnical system fails if its hardware fails, its program crashes, or if users can't figure out or refuse to use it. So if the system fails to perform, it won't survive. Contact tracing apps are a sociotechnology and, as such, they can fail also for social reasons. These social reasons can be seen in motion in the Web which offer us a larger scale view of what the public reactance to other sociotechnical systems are.

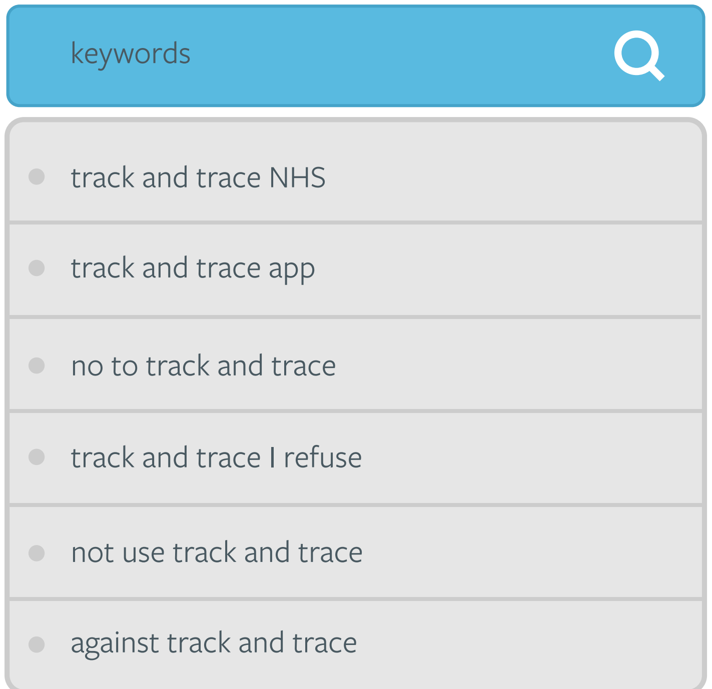
As such, a way to overcome these sociotechnical challenges is understanding the social refusal and resistance to contact tracing technologies. This can be done by drawing from a suitable combination of both the computational tools offered by more technical disciplines (such as the more 'tech' component of Web Science), and the investigative, interpretative and theoretical insights offered by disciplines, such as criminology, with the capacity to uncover and understand complex social patterns and individual trends – a necessary step to shed light on resistance to adoption of contact tracing apps.

METHODOLOGY: DEVELOPING A NEW INTERDISCIPLINARY APPROACH

With our project, we offer an empirical, methodological and conceptual contribution which combines computational capacities to investigate a large social media dataset expanding over 10 months and qualitative expertise in criminology to offer a new angle to reflect on emerging issues of public trust, governance, and the use of personal data for public good that are at the basis of people's resistance in using tracing apps, but that are unlikely to peter out after discussions on COVID-19 contract tracing apps will fade away. In this section, we summarise the approach followed through a sequence of five main stages.

Developing keywords and hashtag lists

A selection of keywords and hashtags specific to the study was required to direct the Twitter search tool. Snowballing sampling was used to determine what hashtags and keywords retrieved the most data to then be used for final data collection. After analysing which keywords and hashtags retrieved the most tweets, we used the following keywords.



Automatised data collection through a computational tool

Datasets were gathered from Twitter using the Web Data Research Assistant software developed by Les Carr, part of the research team in this study. The software is a browser extension for Chrome and Microsoft Edge that monitors pages that the researcher browses (e.g., social media timelines and database search results) and saves relevant data and metadata as a spreadsheet. This software exported information from the Twitter web app search result page, using the researcher-determined search parameters. Once the information was taken, it was made available in an accessible spreadsheet form, summarising key components (title, contents, date, author, etc.). A further process of data enrichment was undertaken by a separate web service (WDRA Extender[2]) which takes lists of Tweets from WDRA and backfills further details about Twitter accounts (e.g., names, locations, number of followers), performs keyword and n-gram analyses, creates a concordance (or keyword-in-context-style (KWIC) indexes) onto the Twitter corpus, and produces network visualisations of the Twitter account interactions. Using this software, tweets were collected retrospectively, in the period starting with the oldest relevant tweet being published on 6th March 2020 and continuing until 31st December 2020.

Identification of relevant hashtags and keywords in the datasets

We identified relevant tweets post hoc from searches in the Twitter Web app[3]. This search produced a total number of 54,941 tweets (including a total of 4,269 hashtags) tweeted from 38,713 Twitter accounts over the 10 months considered. It is worth noting that, of those accounts, 2,530 were considered 'dormant' (that is, they were used to tweet on any subject less than once per week) and 1,437 were probably automated (as they tweeted more than 50 times per day).

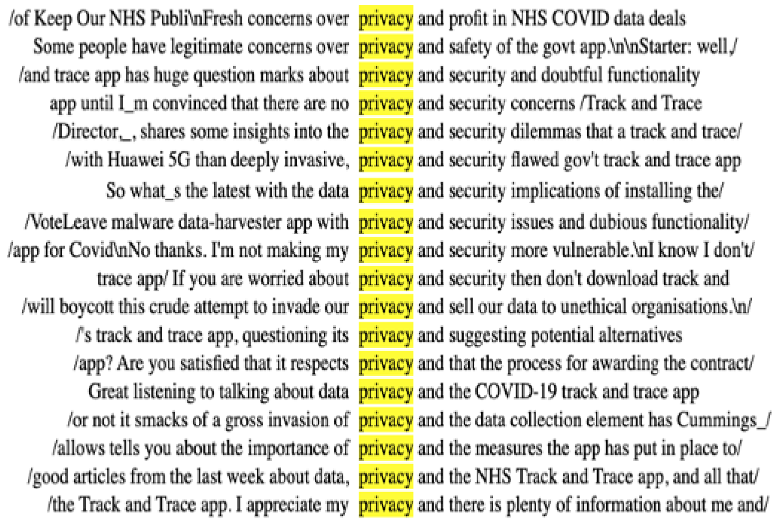
Information extraction and qualitative analyses

Spreadsheets of easily accessible data from WDRA and WDRA Extender were passed to the criminologists for manual analysis. To detect content within the dataset that was both relevant to the keywords and indicative of suspect discourses, practical and effective methods of information extraction were required.

As detailed below, the researchers used: (a) the ***concordance tool*** to understand the context in which the words emerged as useful to then identify the themes relevant for our study that were used; (b) sentiment analysis with ***n-grams*** (that is, phrases of 2 or more words) to expand and refine their conceptualizations, until thematic saturation was reached (approximately after 800-1,000 words per table); (c) a ***conceptual map*** to highlight the main themes and how they interconnected; and (d) ***social network analysis*** to extract information that would allow an understanding of the interaction between the accounts and identify the drivers behind relevant conversations.

(a) Concordance tool

The concordance tool is a way to present the data highlighting the keyword in its original context (Ross & Rivers, 2018). Figure 1, below, shows an example of the keyword-in-context display as a concordance tool for the word 'privacy'.



/of Keep Our NHS Publi\nFresh concerns over	privacy	and profit in NHS COVID data deals
Some people have legitimate concerns over	privacy	and safety of the govt app.\n\nStarter: well./
/and trace app has huge question marks about	privacy	and security and doubtful functionality
app until I_m convinced that there are no	privacy	and security concerns /Track and Trace
/Director,_, shares some insights into the	privacy	and security dilemmas that a track and trace/
/with Huawei 5G than deeply invasive,	privacy	and security flawed gov't track and trace app
So what_s the latest with the data	privacy	and security implications of installing the/
/VoteLeave malware data-harvester app with	privacy	and security issues and dubious functionality/
/app for Covid\nNo thanks. I'm not making my	privacy	and security more vulnerable.\nI know I don't/
trace app/ If you are worried about	privacy	and security then don't download track and
/will boycott this crude attempt to invade our	privacy	and sell our data to unethical organisations.\n/
/'s track and trace app, questioning its	privacy	and suggesting potential alternatives
/app? Are you satisfied that it respects	privacy	and that the process for awarding the contract/
Great listening to talking about data	privacy	and the COVID-19 track and trace app
/or not it smacks of a gross invasion of	privacy	and the data collection element has Cummings_/
/allows tells you about the importance of	privacy	and the measures the app has put in place to/
/good articles from the last week about data,	privacy	and the NHS Track and Trace app, and all that/
/the Track and Trace app. I appreciate my	privacy	and there is plenty of information about me and/

Fig 1 Keyword-in-context display for the word 'privacy'

[2]Available at https://bit.ly/WDRAx
[3]See https://twitter.com/search-advanced

(b) Sentiment analysis with n-grams

Sentiment analysis of free-text documents is a common task in the field of text mining. In sentiment analysis predefined sentiment labels, such as 'positive' or 'negative' are assigned to text documents and N-grams of texts are extensively used in text mining and natural language processing tasks. They are a set of co-occurring words within a given window and when computing the n-grams you typically move one word forward. For the purpose of this study, we used n-grams for semantic analysis, in order to draw meaning from the text.

(c) Conceptual maps

To facilitate a developing conceptualisation of the data, notes were individually taken and then shared, discussed and integrated into the following qualitative conceptual map [see Figure 2 below, which indicates only the main connections identified for clarity purposes] highlighting the main themes and how they are connected.

Qualitative checks for bias minimisation

We evaluated our approach using a triangulation of computational methods and qualitative methods.

To measure the performance of both information extraction techniques, the relevant tweets and the theme identified were iteratively compared by research team members across disciplines.

The qualitative analysis was carried out by the criminologists with insights into how the Web Science tools work in order to further understand the mechanisms of social analysis in Twitter. With the help of a computer scientist team member, the tools were chosen according to the goals of the study.

(d) Social Network Analysis

The aim of social network analysis is to understand a community by mapping the relationships that connect them as a network, and then trying to draw out key individuals, groups within the network ('components'), and/or associations between the individuals. A network is simply a number of points (or 'nodes') that are connected by links (or 'arcs'). Generally, in social network analysis, the nodes are people and the links are any social connection between them – for example, friendship, marital/family ties, or financial ties.

In our study we used social network analysis to map a conversational network obtained by connecting two accounts where one replies to or mentions the other. The network was plotted in Gephi using the ForceAtlas network layout. The algorithm for this layout pulls strongly connected nodes together and pushes weakly connected nodes apart. This way we were able to see which accounts are highly interconnected or being interacted with and which were not. This would allow us to understand the conversation drivers or the type of social media actor setting the tone in the conversations observed.

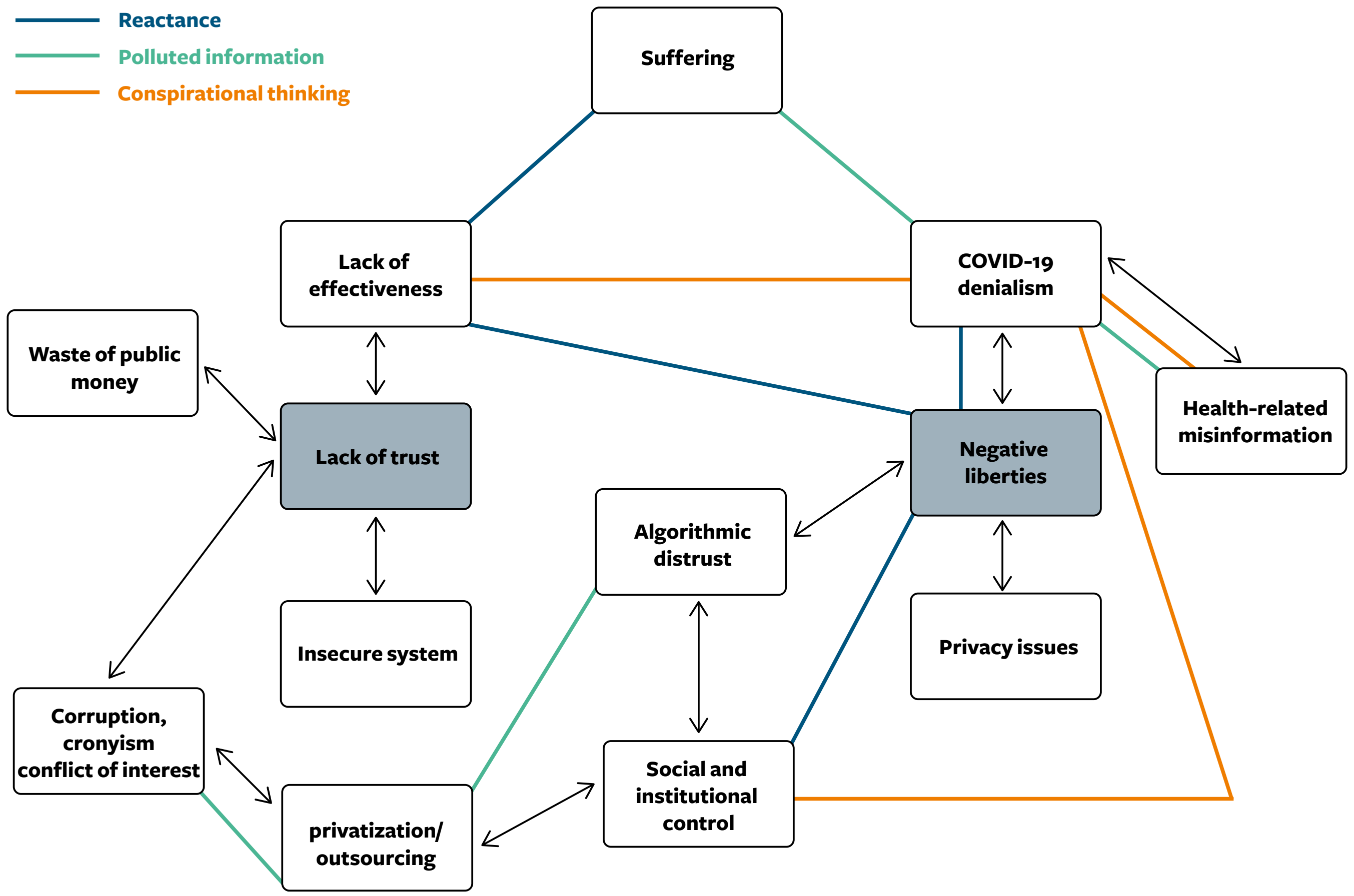


Fig 2 Qualitative conceptual map

THE NETWORKS OBSERVED – MAIN RESULTS

Following our Social Network Analysis of the conversational networks, we were able to identify three main parts in the network: an outer ring, a middle ring and a strongly connected central core. The outer ring consisted of slightly over two thirds of the accounts we gathered information from, followed by the central core with over a quarter, and finally the middle ring with the remaining accounts as seen in Figure 3 below.

Most of the users are found outside the core cluster, as 'isolated individuals' or 'small groups' in the network since they are not involved in any collective conversation around the topic of resistance. However, their opinions were still relevant to our overall analysis since their tweets add up to the themes around resistance.

In order to further understand the sociocultural and other key drivers at play in the network of tweets, we wanted to identify who were the conversation drivers in the network which were represented by the central core of the network as we can see in Figure 4.

Most of the visible conversations were clustered around dominant large clumps of nodes, driven by tweets initiating responses from high status public broadcasters and political organisations; while individually speaking, most of the interaction occurred around people talking to journalists and prominent politicians and the official NHS app.

For further details, please see Lavorgna et al, 2021a.

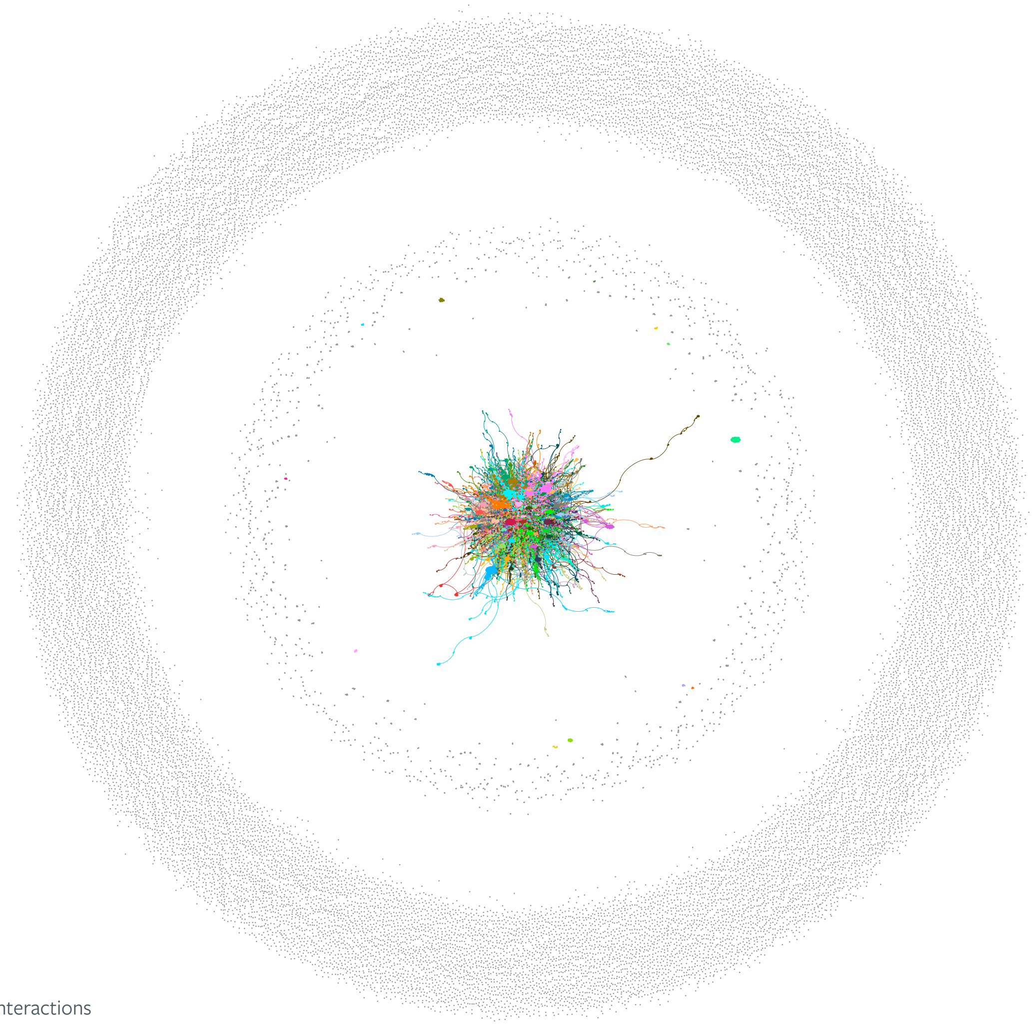


Fig 3 Network of interactions

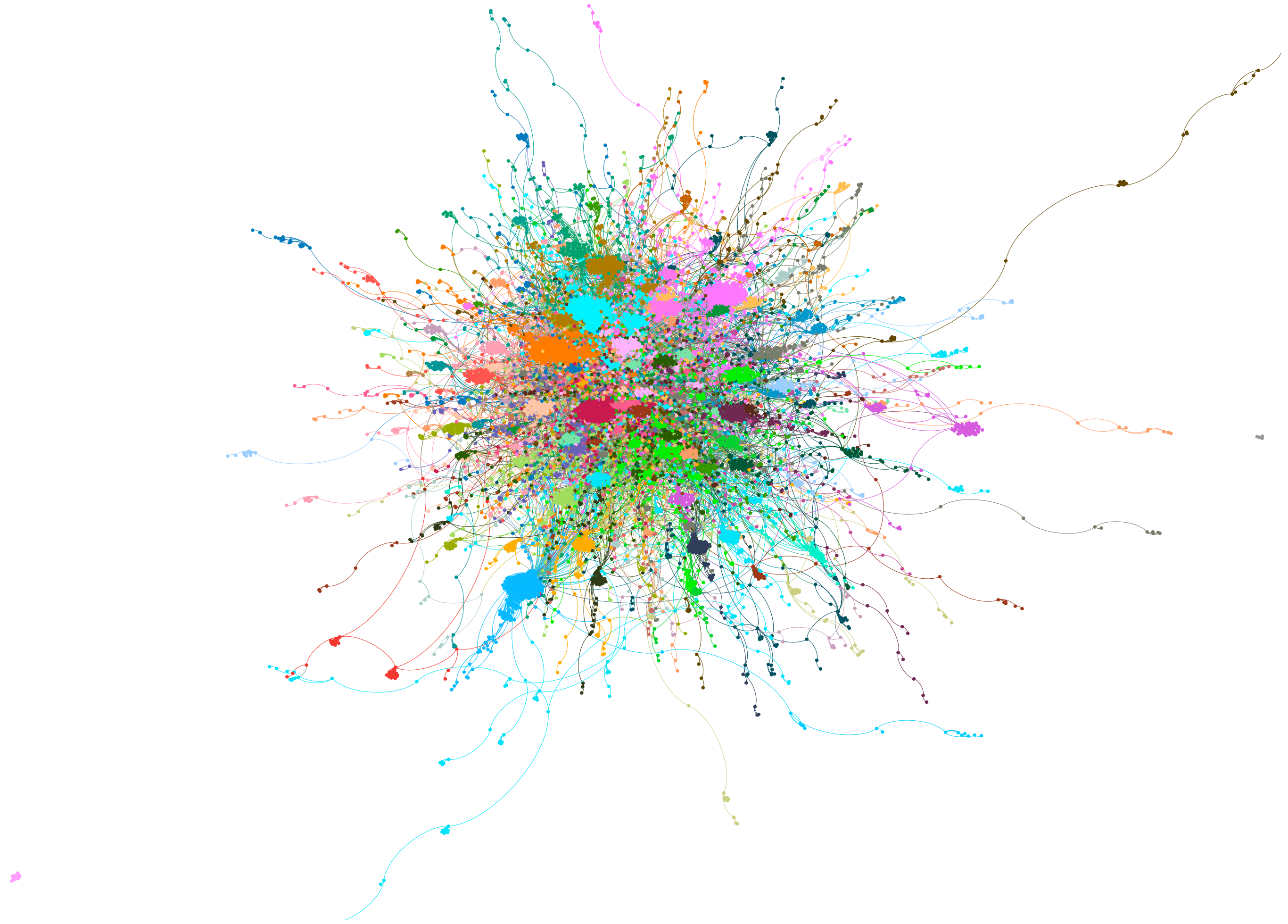


Fig 4 Central core of the network

FRAMES AND MECHANISMS OF RESISTANCE - MAIN RESULTS

Once the networks were observed and we saw how the interactions occurred, following our qualitative analysis into the conversations we identified two main narrative frames and three main mechanisms of resistance in using the NHS contact tracing app. These themes and frames, here briefly addressed, are further discussed in Lavorgna et al., 2021a.

Two of the main narrative frames found in our study were mistrust and negative liberties as the basis of people's resistance in using tracing apps. Mistrust was visible in different ways in the tweets examined. The most visible ones were lack of trust towards the current government; the companies involved in developing the contact tracing app; the effectiveness and security of the app as well as concerns on increasing datafication. At the heart of it, is the perceived incompetence of the actors involved, who are seen as flawed, corrupt, hypocritical, and not accountable for their actions or inactions. Negative liberties (that is, the absence of constraints) of individual agents (rather than the collectives' possibility of acting to realize one's fundamental purposes) were noticed, for instance, in discourses aligned with populist libertarian views, and generally stressing strong opposition to preventive measures such as lockdowns, limitations to travelling and gathering, and the use of masks.

As well as narrative frames informing people's resistance in using the NHS contact tracing app, our study identified three main mechanisms of resistance; polluted information being a key one. Recent research has evidenced how, in the context of medical misinformation, polluted information can lead to serious social harms (Lavorgna, 2021). From our study, it was found that polluted information granted access to a high number of misleading health related information facilitating, for instance, antimask and antivax views as well as questioning social distancing guidelines. Another example of polluted information enabling certain public reactions we found was the propagation of false and misleading information on the role of public companies in the development of the contact tracing app. Many users, for instance, tweeted about a major private company operating in the field of public services, linking it with themes around corruption, conflict of interest and lack of trust. However, that company played no role in the creation of the NHS tracing app nor did it process its data (FullFact, 2020).

Another mechanism of resistance found in our study was conspiratorial thinking. This mechanism posits the idea that there are agents (individuals, groups or organizations) working together and plotting to accomplish menacing objectives (van der Linden, 2013; van der Linden et al., 2020). In our study, conspiratorial thinking was identified as the driving force behind COVID-19 denialism and behind the idea that the app is part of a clandestine plan for mass control. But conspiratorial thinking is not an isolated phenomenon. As a result of various elements already mentioned in this report, such as the role of social media platforms and digital technologies in facilitating high-speed information, environments that facilitate false and fake narratives in terms of popularity and audience engagements (Silverman, 2016); narratives of conspiracy theories and misinformation spread quickly, especially in times of societal crisis such as the COVID-19 pandemic (e.g., De Coninck et al., 2020; Imhoff & Lamberty, 2020; Knuutila et al., 2020)

The third mechanism of resistance identified in our study was reactance, which refers to how people tend to be averse to having restricted freedom or ability to act in a particular way. When this happens, they tend to reject evidence that is perceived as a threat to their ability to act (or do not act) in a certain way (Rosenberg, 2018; Prot, 2019). In general, people believe that they possess certain freedoms to engage in so-called free behaviours, yet there are times when they cannot or at least feel they cannot, do so; and it serves as a motivator to restore one's freedom (Steindl et al., 2015).

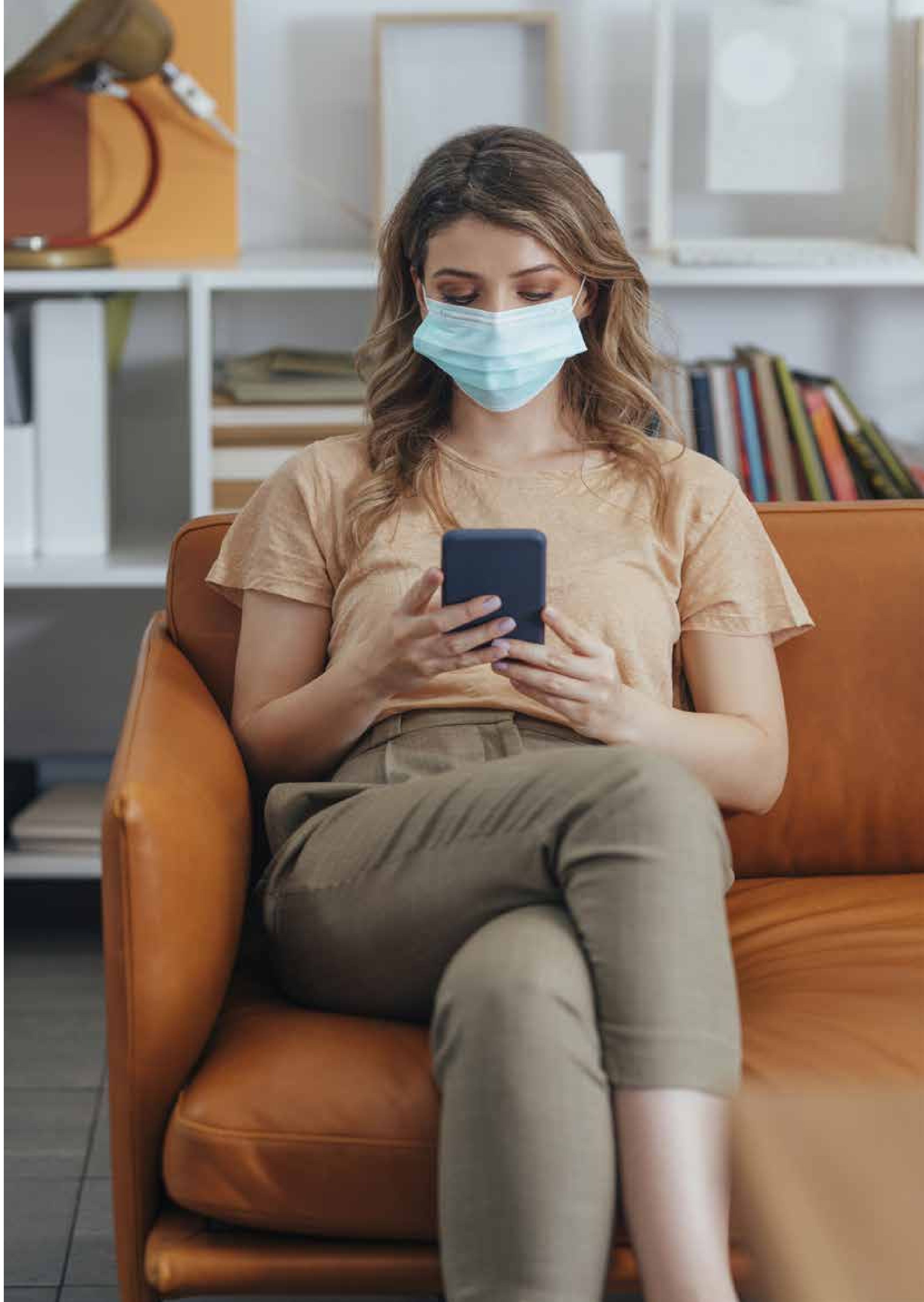
INSIGHTS INTO THE ROLE OF SOCIAL NETWORKING SITES INFRASTRUCTURES MECHANISMS THAT AID MISINFORMATION IN PUBLIC DISCOURSE

As we have seen, the COVID-19 pandemic was mirrored by the diffusion of misinformation and conspiracy theories; beliefs that have resulted in substantive, harmful outcomes but which remain largely unstudied (Agley & Xiao, 2021). Because of how their infrastructure work, SNSs have been at the centre of this diffusion (e.g., Reisach, 2021. See also section 1.4 of this Report).

SNSs originally had positive connotations for creative engagement, political participation, cross promotion, and the ability to allow even the small citizen to have a loud voice (e.g., Dwyer & Martin, 2017; Reisach, 2021). However, SNSs have a business model which is driven by commercial needs that favour advertisers and consumers (Bardoel & d'Haenens, 2004) and vast opinion power (Helberger, 2020). For instance, research shows that all web technology platforms, such as SNSs and news recommendations systems, have a strong popularity bias (Nikolov et al., 2019), meaning that our consumption of online information is mediated by filtering, ranking and recommendations algorithms that introduce unintentional biases as they attempt to deliver relevant and engaging content. Dependence on engagement metrics may also make us vulnerable to manipulation by orchestrated campaigns and social bots (Ratkiewicz et al., 2011; Ferrara et al., 2016). Furthermore exposure to news through the filter of the social network of like-minded individuals may bias our attention toward information that we are already likely to know or agree with (Conove et al., 2011; Pariser, 2011; Sunstein, 2017). A quantitative study (Nikolov et al., 2019) found that social media and search engines tend to be the primary channels through which users consume new information where social media tends to exhibit more homogeneity bias (that is, the tendency of a platform to expose users to information from a narrow set of sources) and popularity bias (that is, the tendency of a platform to expose users to information from popular sources) compared to search engines. This is consistent with previous research findings (Nikolov et al., 2015) that social media may contribute to the emergence of 'social bubbles'.

In order to counter misinformation and societal polarization, a responsibility-based approach for social media platforms (e.g., Reisach, 2021) should be welcomed. There are ongoing policy initiatives, especially in the European region, to impose greater social responsibility on social media platforms with aims at combating disinformation and appeals to install a transparent and consistent moderation of disinformation (European Commission, 2020a,b). Also, better data (and metadata) availability for researcher could help to better understand the social and computational mechanisms at the basis of misinformation in public discourse, and hence allow us to think about better preventive and contrasting strategies (e.g., Pasquetto et al., 2020).

The use of analysis, clustering, tracking, prediction and recommendations systems are powerful tools for supporting decision making; since these have been growing exponentially with the aid of machine learning algorithms, so has their significance in citizens' political decision making (Reisach, 2021). To ensure the benefits of social media platforms and avoid social harms depends heavily on the shared interests between social media platforms, their potential benefits and risks, and the transparency of their infrastructures' processes. Sociotechnical thinking and research can support finding practical and effective ways to find appropriate yet feasible balances in this complex process.



TRACING APPS AS ALTERNATIVE CONTROL INTERVENTIONS

Resistance to Public Health Initiatives (PHIs) is a longstanding problem and this project demonstrates that it is a key factor in the COVID-19 pandemic, with resistance amplified by populist actors participating publicly on SNSs. Clearly, understanding the multifaceted nature of this problem is vital to ensure the effectiveness of PHIs. An area that deserves further theoretical and empirical scrutiny relates to the cultures of such resistance, and this is the problem on which the project will focus in its next steps. It will explore public discourses on SNSs, specifically Twitter, Facebook, and Instagram, to understand key cultural expressions of resistance to PHIs and develop remedial digital resources. Building on insights from cultural criminological studies of transgressive acts (see, e.g., Ferrell et al., 2015), of which resistance to PHIs represents an example, the project – as it moves forward – will proceed on the basis that, to further understand the sources and patterns of resistant practices on SNSs and develop remedies, it is important to study and understand the cultures of such resistance.

Perspectives in cultural criminology contend that culture does not exist in a vacuum; it intersects with structural inequalities of age, gender, race, class, religion, and other social categories which provide the context in which culture is expressed (e.g., Presdee, 2004). As such, to develop a deeper understanding of the problem of resistance to PHIs in general, including but not solely the COVID-19 initiatives, the future project will explore and understand intersections of the cultures and structures of such resistance. The research will integrate expertise from social, health, and computer sciences to develop a broad, interdisciplinary approach in its analysis of these issues. It will generate new insights and digital resources that relevant services and other authorities can integrate into public health systems to detect and counter such practices whilst mobilising public trust and support.

Although the current study usefully reveals several forms of resistance to PHIs expressed for example through information pollution (e.g., via misinformation practices), the future project takes this step further by offering unique insight into the cultures of such resistance and how they are expressed in digital environments, using shared, symbolic practices. Studying the sources and patterns of such cultures is crucial for effective remediation. The study's outputs will therefore support efforts to counter resistance, improve public health communications, and promote sustainable public trust in PHIs.

REFERENCES

Abeler J, Altmann S, Milsom L, Toussaert S & Zillessen H (2020). Support in the UK for App-based Contact Tracing of COVID-19. Department of Economic, University of Oxford, United Kingdom.

Ada Lovelace Institute (2020) Exit Through the App Store? Retrieved from https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-RapidEvidence-Review-Exit-through-the-App-Store-April-2020-2.pdf.

Agley J & Xiao Y (2021) Misinformation about COVID-19: evidence for differential latent profiles and a strong association with trust in science. BMC Public Health 21, 89.

Akinbi A, Forshaw M & Blinkhorn V (2021) Contact tracing apps for the COVID-19 pandemic: a systematic literature review of challenges and future directions for neo-liberal societies. Health Inf Sci Syst 9(18).

Alrajebah N, Carr L & Tiropanis T (2018) The platform effect: Analysing user activity on tumblr. In Internet Science: INSCI 2018. vol. 11193 LNCS, Springer, 154-168.

Alrajebah N, Carr L, Luczak-Roesch M & Tiropanis T (2017) Deconstructing diffusion on Tumblr: structural and temporal aspects. In WebSci '17 Proceedings of the 2017 ACM on Web Science Conference. ACM, 319-328.

Alsdurf H, Bengio Y, Deleu T, Gupta P, Ippolito D, Janda R, Jarvie M, Kolody T, Krastev S, Maharaj T, Obryk R, Pilat D, Pisano V, Prud'homme B, Qu M, Rahaman N, Rish I, Rousseau J, Sharma A, Struck B, Tang J, Weiß M & Yu Y (2020) COVI White Paper. ArXiv, abs/2005.08502.

Angwin J & Larson J (2016) Bias in criminal risk scores is mathematically inevitable, researchers say. Retrieved from https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say.

Arpey NC, Gaglioti AH. & Rosenbaum ME (2017) How socioeconomic status affects patient perceptions of health care: A qualitative study. Journal of Primary Care & Community Health 8(3), 169-175.

Auernhammer J (2020) Human-centered AI: the role of Human-centered Design Research in the development of AI. In Boess S, Cheung M & Cain R (eds.) Synergy - DRS International Conference.

Bardoel J & d'Haenens L (2004) Media responsibility and accountability: new conceptualizations and practices. Communications 29, 5-25.

Bellini R, Dell N, Whitty M, Bhattacharya D, Wall D & Briggs P (2020) Crime and/or Punishment: Joining the Dots between Crime, Legality and HCI. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1-8.

Blaynee J, Kreps DM, Kutar M & Griffiths M (2016) Collaborative HCI and UX: longitudinal diary studies as a means of uncovering barriers to digital adoption. BCS HCI.

Boniface M, Carmichael L, Hall W, Pickering B, Stalla-Bourdillon S & Taylor S (2020) A blueprint for a social data foundation: Accelerating trustworthy and collaborative data sharing for health and social care transformation. Web Science Institute White Papers, 4. University of Southampton.

Budd J, Miller BS, Manning EM et al. (2020) Digital technologies in the public-health response to COVID-19. Nat Med 26, 1183-1192.

Carroll J (1997) Human-computer interaction: psychology as a science of design. Annual Review of Psychology 48, 61-83.

Conover M, Ratkiewicz J, Francisco M, Gonçalves B, Flammini A & Menczer F (2011) Political polarization on Twitter. In Proceedings of the Fifth International AAAI Conference on weblogs and Social Media (ICWSM).

De Coninck D, d'Haenens L & Matthijs K (2020) Forgotten key players in public health: News media as agents of information and persuasion during the COVID-19 pandemic. Public Health 183, 65-66.

Dombrowski L, Harmon E & Fox S (2016) Social Justice-Oriented Interaction Design: Outlining Key Design Strategies and Commitments. Proceedings of the 2016 ACM Conference on Designing Interactive Systems, ACM, 656-671.

Douglas M, Katikireddi SV, Taulbut M, McKee M & McCartney G (2020) Mitigating the wider health effects of covid-19 pandemic response. BMJ 369:m1557.

Dwyer T & Martin F (2017) Sharing News Online. Digital Journalism, 5(8), 1080-1100.

European Commission (2020a) Digitale Wirtschaft – Kommission strebt mehr Wettbewerb im Internet an (Digital Economy – Commission strives for more competition in the internet). Vertretung in Germany. In: Representation in Germany: EU-News no.11/2020, 1-2.

European Commission (2020b) European Commission European data strategy. Strategy priorities 2019–2024: a Europe fit for the digital age. Brussels.

Farries E (2020). Covid-tracing app may be ineffective and invasive of privacy Government must be transparent to avoid unintended consequences. The Irish Times, 5 May.

Farronato C, Iansiti M, Bartosiak M, Denicolai S, Ferretti L & Fontana R (2020). How to Get People to Actually Use Contact-Tracing Apps. Harvard Business Review.

Felzmann H, Fosch-Villaronga E, Lutz C et al. (2020) Towards transparency by design for Artificial Intelligence. Sci Eng Ethics 26, 3333-3361.

Ferrara E, Varol O, Davis C, Menczer F & Flammini A (2016) The rise of social bots. Communications of the ACM 59(7), 96-104.

Ferreira A (2016) Universal UX Design: Building Multicultural User Experience. Massachusetts: Morgan Kaufmann.

Ferrell J, Hayward K & Young J (2015) Cultural criminology: An invitation (2nd ed.). Los Angeles, CA: SAGE.

Fitriani (2020) COVID-19 Apps: Fear of Tyranny by Data. The Jakarta Post. Retrieved from https://www.thejakartapost.com/academia/2020/06/22/covid-19-apps-fear-of-tyranny-by-data.html.

FullFact (2020) Serco didn't build and does not run the NHS Test and Trace app. Retrieved from https://fullfact.org/health/Serco-test-and-trace/.

Gannoni A, Willis M, Taylor E & Lee M (2017) Surveillance technologies and crime control: understanding police detainees' perspectives on police body-worn video (BWV) and CCTV cameras (CRG 31/14-15). Australia: Criminology Research Advisory Council.

Garrett P, White J, Little D, Perfors A, Kashima Y, Lewandowsky S & Dennis S (2020). A Representative Sample of Australian Participant's Attitudes Towards the COVIDSafe App. Complex Human Data Hub, School of Psychological Sciences, The University of Melbourne, Australia.

Gianfrancesco MA, Tamang S, Yazdany J & Schmajuk G (2018) Potential biases in machine learning algorithms using electronic health record data. JAMA Internal Medicine 178(11), 1544-1547.

González-Padilla DA & Tortolero-Blanco L (2020) Social media influence in the COVID-19 Pandemic. Int Braz J Urol. 46, 120-124.

Greenhalgh T & Papoutsi C (2019) Spreading and Scaling Up Innovation and Improvement. The BMJ, 365, l2068.

Halford S (2020) On multidisciplinarity, the making of socio-digital futures, their challenges and opportunities. World Podcasts. Retrieved from https://worldpodcasts.com/susan-halford-on-multidisciplinarity-the-making-of-socio-digital-futures-their-challenges-and-opportunities/.

Hall W & Tiropanis T (2012) Web evolution and Web Science. Computer Networks 56(18), 3859-3865.

Helberger N (2020) The political power of platforms: how current attempts to regulate misinformation amplify opinion power. Digital Journalism 8(6), 842-854.

Hendler J, Shadbolt N, Hall W, Berners-Lee T & Weitzner D (2008) Web science: an interdisciplinary approach to understanding the Web. Communications of the ACM 51(7), 60-69.

Hern A (2020) Millions risk being locked out of Covid-19 contact tracing app. The Guardian, May 15.

Hogan K, Macedo B, Macha V, Barman A & Jiang X (2021) Contact Tracing Apps: Lessons Learned on Privacy, Autonomy, and the Need for Detailed and Thoughtful Implementation. JMIR medical informatics 9(7), e27449.

Hooper CJ & Dix A (2013) Web science and human-computer interaction: forming a mutually supportive relationship. Interactions 20(3), 52-57.

Imhoff R & Lamberty P (2020) A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. Social Psychological and Personality Science 11(8), 1110-1118.

Ireton C, Posetti J (2018) Journalism, 'fake news' & disinformation: handbook for journalism education and training. Unesco.

Kim L, Fast SM & Markuzon N (2019) Incorporating media data into a model of infectious disease transmission. PLoS ONE 14.

Knuutila A, Herasimenka A, Au H, Bright J, Nielsen R & Howard PN (2020) COVID-related misinformation on Youtube. Oxford Institute University.

Koumchatzky N & Andryeyev A (2017) Using Deep Learning at Scale in Twitter's Timelines. Retrieved from: https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.

Kuziemski M & Misuraca G (2020) AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications policy 44(6), 101976.

Lavorgna A & Carr L (2021) Tweets and quacks: network and content analyses of providers of non-science-based anti-cancer treatments and their supporters on Twitter. Sage Open.

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

Lavorgna A & Myles H (2021) Science denial and medical misinformation in pandemic times: a psycho-criminological analysis. European Journal of Criminology.

Lavorgna A & Ugwudike P (2021) The datafication revolution in criminal justice: an empirical exploration of frames portraying data-driven technologies for crime prevention and control. Big Data & Society.

Lavorgna A, Carr L, Ugwudike P, Sanchez Benitez Y & Rekha SG (2021a) To app or not to app? Understanding public resistance in using COVID-19 digital contact tracing. Law, Technology & Humans.

Lavorgna A, Carr L & Kingdon A (2021) To wear or not to wear? Unpacking the #NoMask discourses and conversations on Twitter..

Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F & Schudson M (2018) The science of fake news. Science 359, 1094-1096.

Lee J & Choi Y (2018) Informed public against false rumor in the social media era: focusing on social media dependency. Telematics and Informatics 35(5), 1071-1081.

Lewis D (2021) Contact-tracing apps help reduce COVID infections, data suggest. Nature 591(7848), 18-19.

Llewellyn S (2020) COVID-19: How to be careful with trust and expertise on social media. British Medical Journal 368, m1160.

Marjanovic S, Altenhofer M, Hocking L, Chataway J & Ling T (2020) Innovating for improved healthcare: Sociotechnical and innovation systems perspectives and lessons from the NHS. Science and Public Policy 47(2), 283-297.

Masel J, Shilen A, Helming B, Rutschman J, Windham G, Pogreba-Brown K & Ernst K (2021) Quantifying meaningful adoption of a SARS-CoV-2 exposure notification app on the campus of the University of Arizona. MedRxiv, 2021.02.02.21251022.

Milan S (2020) Techno-solutionism and the standard human in the making of the COVID-19 pandemic. Big Data & Society.

Nellis M (2015) Standards and ethics in electronic monitoring. Handbook for professionals responsible for the establishment and the use of electronic monitoring. Retrieved from https://rm.coe.int/handbook-standards-ethics-in-electronic-monitoring-eng/16806ab9b0.

NHS Digital (2017) NICE Technology Appraisals in the NHS in England (Innovation Scorecard): To December 2016. Leeds: Health & Social Care Information Centre. Retrieved from: https://files.digital.nhs.uk/publication/a/e/nice-tech-apps-eng-jul17-inno-scor-rep.pdf.

Nikolov D, Oliveira DFM, Flammini A & Menczer F (2015) Measuring online social bubbles. PeerJ Computer Science 1, e38.

Nikolov D, Lalmas M, Flammini A & Menczer F (2019) Quantifying biases in online information exposure. Journal of the Association for Information Science and Technology 70, 218-229.

Okamoto M & Fujita T (2020) Contact tracing apps have a privacy problem. This new framework could help. World Economic Forum. Retrieved from https://www.weforum.org/agenda/2020/08/contact-tracing-apps-privacy-framework-appa-data-governance/.

Pagliari C (2020) The ethics and value of contact tracing apps: international insights and implications for Scotland's COVID-19 response. Journal of Global Health 10(2).

Pariser E (2011) The filter bubble: how the new personalized web is changing what we read and how we think. New York: Penguin.

Pope C & Turnbull J (2017) Using the concept of hubots to understand the work entailed in using digital technologies in healthcare. Journal of Health Organization and Management 31(5), 556-566.

Presdee M (2004) Cultural Criminology: the long and winding road. Theoretical Criminology 8(3), 275-285.

Privacy International (2021) Tracking the Global Response to COVID-19. Retrieved from: https://privacyinternational.org/examples/tracking-global-response-covid-19.

Prot S & Anderson CA (2019) Science denial. Psychological processes underlying denial of science-based medical practices. In Lavorgna A & Di Ronco A (eds.). Medical Misinformation and Social Harm in Non-Science-Based Health Practices: A Multidisciplinary Perspective. Routledge.

Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Flammini A & Menczer F (2011) Detecting and tracking political abuse in social media. In Proceedings of the Fifth International AAAI Conference on weblogs and Social Media (ICWSM).

Reisach U (2021) The responsibility of social media in times of societal and political manipulation. European Journal of Operational Research 291(3), 906-917.

Rodríguez P, Graña S, Alvarez-León EE et al. (2021) A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. Nature Communications 12, 58.

Rosenberg BD & Siegel JT (2018) A 50-year review of psychological reactance theory: do not read this article. Motivation Science 4(4), 281-300.

Ross AS & Rivers DJ (2018) Discursive deflection: accusation of "fake news" and the spread of mis- and disinformation in the tweets of President Trump. Social Media + Society, 4(2).

Salvi C, Iannello P, Cancer A, McClay M, Rago S, Dunsmoor JE & Antonietti A (2021) Going viral: how fear, socio-cognitive polarization and problem-solving influence fake news detection and proliferation during COVID-19 pandemic. Frontiers in Communication, 5, 127.

Sanchez Benitez Y (2021) Algorithmic ecologies of justice: Using computational social science methods to co-design tools of resistance, resilience and care with communities. 13th ACM Web Science Conference 2021 (WebSci '21). Association for Computing Machinery, New York, USA, 149-150.

Sánchez-Monedero J, Dencik L & Edwards L (2020) What does it mean to 'solve' the problem of discrimination in hiring? social, technical and legal perspectives from the UK on automated hiring systems. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 458-468.

Shin D (2020). User Perceptions of Algorithmic Decisions in the Personalized AI System:Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. Journal of Broadcasting & Electronic Media.

Shin D, Zhong B & Biocca FA (2020) Beyond user experience: What constitutes algorithmic experiences? International Journal of Information Management.

Silverman C (2016) This analysis shows how viral fake election news stories outperformed real news on Facebook. New York: Buzzfeed.

Stanney KM, Smith MJ, Carayon P, Salvendy G (2001) Human-computer interaction. In Salvendy G (ed.) Handbook of Industrial Engineering, 3rd edn. John Wiley.

Steindl C, Jonas E, Sittenthaler S, Traut-Mattausch E & Greenberg J (2015) Understanding Psychological Reactance: New Developments and Findings. Zeitschrift fur Psychologie 223(4), 205-214.

Sunstein CR (2017) #Republic: Divided democracy in the age of social media. Princeton, NJ: Princeton University Press.

Sweeney Y (2020) Tracking the debate on COVID-19 surveillance tools. Nat Mach Intell 2, 301-304.

Tufekci Z (2017) Twitter and Tear Gas: The Power and Fragility of Networked Protest. Yale University Press.

Ugwudike P (2020) Digital prediction technologies in the justice system: The implications of a 'race-neutral' agenda. Theoretical Criminology 24(3), 482-501.

Ugwudike P (2021) Data-driven technologies in justice systems: intersections of power, data configurations, and knowledge production. In Lavorgna A & Holt TJ (eds) Researching Cybercrimes. Methodologies, ethics and critical approaches. Palgrave Macmillan.

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

van der Linden S (2013). What a hoax: Why people believe in conspiracy theories. Scientific American Mind 24(4), 41-43.

van der Linden S, Panagopoulos C, Azevedo F & Jost JT (2021) The Paranoid Style in American Politics Revisited: An Ideological Asymmetry in Conspiratorial Thinking. Political Psychology 42(1), 23-51.

van Dijck J & Alinejad D (2020) Social Media and Trust in Scientific Expertise: Debating the Covid-19 Pandemic in The Netherlands. Social Media + Society.

Verma N & Dombrowski L (2018) Confronting Social Criticisms: Challenges when Adopting Data-Driven Policing Strategies. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, ACM, 1-13.

Waterson P, Robertson MM, Cooke NJ, Militello L, Roth E & Stanton NA (2015) Defining the methodological challenges and opportunities for an effective science of sociotechnical systems and safety, Ergonomics 58(4), 565-599.

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

Whitworth B (2009) The social requirements of technical systems. In Whitworth B & de Moor A (eds.) Handbook of Research on Socio-Technical Design and Social Networking Systems. IG Global.

Whitworth B & Ahmad A (2013) Socio-Technical System Design. The Encyclopedia of Human-Computer Interaction, 2nd Ed. The Interaction-Design.org Foundation, Aarhus.

Wymant C, Ferretti L, Tsallis D et al. (2021) The epidemiological impact of the NHS COVID-19 app. Nature 594, 408-412.

Završnik A (2020) Criminal justice, artificial intelligence systems, and human rights. ERA Forum 20, 567-583.

Dr Anita Lavorgna (PI), Associate Professor in Criminology
Sociology, Social Policy & Criminology, University of Southampton
a.lavorgna@soton.ac.uk
@anitalavorgna

Dr Pamela Ugwudike, Associate Professor in Criminology
Sociology, Social Policy & Criminology, University of Southampton
p.ugwudike@soton.ac.uk
@PamelaUgwudike

Ms Yadira Sanchez Benitez, PhD student
Sociology, Social Policy & Criminology, University of Southampton
y.sanchezbenitez@soton.ac.uk
@yadira_sz

Prof Les Carr, Head of Web and Internet Science Group
Web Science Institute, University of Southampton
lac@ecs.soton.ac.uk
@lescarr