LSE Media and Communications

Media@LSE Working Paper Series
Editors: Bart Cammaerts, Nick Anstead and Richard Stupart

The Uncertain Decorum of Online Identification

A Study in Qualitative Interviews

Samuel DiBella

Published by Media@LSE, London School of Economics and Political Science ("LSE"), Houghton Street, London WC2A 2AE. The LSE is a School of the University of London. It is a Charity and is incorporated in England as a company limited by guarantee under the Companies Act (Reg number 70527).

ABSTRACT

This study explores the ethics and motivations of online identification—how and why people collect and publish identifying information about others online. In seven interviews, activists, Internet users, advocates, and journalists were asked about their investigative practice and how they viewed the ethics of deanonymization. Using ethnographic interviewing techniques and a thematic analysis inspired by grounded theory, I describe respondents' investigations and compare them to existing theories in surveillance studies, online anonymity, and digital vigilantism.

Respondents often struggled with making their work accessible and impactful in an ethical manner. They obfuscated irrelevant information that might incite online harassment and took care in who they collaborated with. The respondents also debated what to do when people misinterpreted their work or thought that they had acted unjustly. The precautions they incorporated into their publications are examples of how people navigate online ethics when there isn't a clear standard for moral decisions.

Ultimately, the interview results did not follow models of digital vigilantism and doxxing, and I caution against using those terms to apply to cases like those described in this study. I also make suggestions for how these results could augment theoretical models of anonymity, particularly how respondents' investigative techniques and backgrounds lead them to different moral commitments.

1 INTRODUCTION

> *I hated the idea that he knew everything about me while I knew little or nothing of him. I felt like someone who is blind and knows that he is being observed by the very people he would like to spy on in every detail.*
> — Elena Ferrante, *The Days of Abandonment*

Debates about digital privacy often center around the handling of data by powerful actors—companies and countries. In parallel, researchers have been discussing the extent of online harassment, especially in massive, organized campaigns, and the mental-health harms they cause (Lenhart *et al.*, 2016). But collecting identifying information is also within grasp of the everyday Internet user. A tide of digital information has made loss of anonymity an ordinary occurrence on the Internet. With social media data, public records, and tools like satellite imagery and reverse-image searching, people can easily trace the online activities of friends, acquaintances, coworkers, or strangers.

These are new ways of learning about people that depend on unofficial sources and the unknowing complicity of others (whoever shared the information initially). In the current iteration of Internet life, there's little vocabulary for describing group privacy responsibilities or for deciding what is alright to do with someone else's information. And the differing vulnerabilities people have online, based on their gender, race, citizenship status, or economic class, make it hard to assume what will or won't harm another person (Gangadharan, 2017). Asking for permission to post photos to social media is common now, even as it has become normal to look up the personal or professional social media accounts of someone you've just met. The norms for using identifying information are still evolving as Internet communities adapt digital investigative techniques to their own ends.

This study arose out of my activist work on digital privacy education in the United States. I wanted to learn how widely available personal information changes the way people relate to one another. Rather than focus on institutional attempts to preserve privacy and protect data online, I decided to explore the messy negotiations and ethical decisions that individuals make when handling data. This study answers some of my initial questions by uncovering the asymmetrical relationship between those who acquire information and those they collect information about.

I'll begin with a literature review of the theories that guided those initial questions and move on to a description of how I designed the interviews. I'll then present the results of the interviews and try to answer my research questions using only information provided by respondents. By doing so, I follow the traditions of grounded theory by giving as clear a description of the data as possible before comparing them to abstract theory. I'll finish by connecting those results to theories of anonymity, digital vigilantism, and surveillance studies and making some final conclusions and recommendations for future research.

2 THEORETICAL BACKGROUND & RESEARCH QUESTIONS

In this literature review, I'll discuss several bodies of literature that relate to the investigation and identification of people online. First, media archaeology on files illuminates how the format of stored files and archives affects investigations, as well as how several of the interviewees contributed to or created archival databases. Drawing from surveillance studies, I'll talk about the kinds of vigilant watching that constitute surveillance, lateral surveillance and sousveillance. Studies in computer-mediated communication and digital vigilantism will provide context about the norms of Internet culture and how Internet communities pursue justice or address social violations. Finally, I'll use theories of privacy and anonymity to flesh out how investigators viewed the ethics of removing someone's anonymity online and how they tried to protect their own.

2.1 The Media of Files and Documents

The practical study of communication material has existed in fields like diplomatics and paleography for hundreds of years (Duranti, 1989). Works like Friedrich Kittler's (1990) *Discourse Networks, 1800/1900*, however, encouraged media theorists to pay closer attention to genres of communicating and recording. The genre of a piece of writing and the way it is stored, organized, and circulated for use, helps to us understand the meaning of the writing and its historical importance.

Integral to that project, Cornelia Vismann's (2008) *Files* traces the establishment of German bureaucracy through the institutions that handled its records. Vismann starts with the oral tradition of Roman law and how the informal notes and personal files of praetors (administrative magistrates) eventually became codified and entombed in state archives. She argues that the stratified nature of the Roman Empire 'was an effect of archiving, that is, of immobilizing files' (Vismann, 2008: 59).

Copy after copy of files spread and become meaningless in archives unless readers can understand their original context. Vismann and others have shown how mechanisms for tracking and verifying different versions of files, like registers, memos, and chanceries, made modern bureaucracy possible, and even desirable (Beniger, 1997; Guillory, 2004). As a result, the current model of a state *depends* on written record; it couldn't function without the endless and confusing proliferation of files.

Other writers have extended Vismann's approach to describe other countries and more contemporary means of circulating writing, as Lisa Gitelman does in her work on copy blanks and xeroxing in the United States. Gitelman (2014: 2) also spends more time on documents—papers that serve as evidence of the very claim they make. The genre of documents and the need for bureaucracies to understand *through* records lead to new forms of public disclosure, as Gitelman shows in her discussion of the leaked US Pentagon Papers in 1967 and the illicit copies of the Unix kernel and user manual released in 1976. Underground publishing circuits like samizdat in Russia (Komaromi, 2012) and zines in the United States often distributed these illicit copies outside of their institutions of origin. The ease of

copying granted by technology like xerox machines and carbon copies also enabled people to build their own personal archives and dossiers (Reichel, 1977).

Modern Internet databases resemble earlier chanceries and archives in that their records usually don't include the context of their creation and information storage, and their data do not circulate freely. Still, file circulation is easy online—copies are easy to transmit and impromptu archives are easy to form. Copies of our information are made constantly by tracking scripts and web scraping tools, and together those copies trace our online behavior (Reigeluth, 2014). Files spread over the Internet resemble Hito Steyerl's (2012; 32) concept of the 'poor image', a file that is no more than a 'copy in motion'. Without the documentation or help of their creators, most databases would be useless references for new readers. Without that context, the lifespan of data's use is actually quite short. Data is ephemeral, despite the intention of its creators (Kallinikos, 2009).

Internet databases can form clear, structured sources of information, but raw data collected en masse is more likely to produce a confusing heap. These 'informational middens' intersect with institutional archives in unforeseen ways. They can illuminate an unnoticed connection or obscure that same connection beneath a deluge of irrelevant data—a problem made worse by broken links, lost content, and software incompatibility (Brunton, 2017: 142).

2.2 Surveillance, Lateral Surveillance, and Sousveillance

In *Discipline & Punish*, Michel Foucault (1995) examined Jeremy Bentham's 'panopticon', a design for a prison arranged around a central tower that prevented prisoners from seeing the guards stationed to watch them. Bentham thought that because prisoners would not know *when* they were surveilled, they would have to internalize prison rules to ensure they weren't acting out of order when a guard looked their way. The panopticon was key for Foucault's examination of penal institutions in the modern era, as methods of discipline spilled out of the judicial system into institutions like hospitals and schools.

Later theorists, however, challenged Foucault's emphasis on disciplinary surveillance. They argued that mass media like television enabled other types of vigilant watching; mass media form a 'synopticon' that allows the many to surveil the few (Mathiesen, 1997). And while surveillance is normally associated with the development of the modern bureaucratic state, other institutions and groups can perform surveillance just as well (Coleman and McCahill, 2011). Building on Deleuze and Guattari's (1987) non-hierarchical networks and Donna Haraway's (1991) theory of the feminist cyborg, who blurs the line between human and non-human, Richard Haggerty and Kevin Ericson (2000) coined the 'surveillant assemblage'. A surveillant assemblage is a collection of human and non-human actors that *together* carry out the watching of others. A guard watching a CCTV monitor is a simple example, but a wandering tourist taking pictures and uploading them to a server is also part of a surveillant assemblage.

The spread of mobile phones, Internet access, and digital cameras in the 21st century made systemic 'sousveillance', or observation from below, possible. Unlike surveillance, which records and profiles those marginalized and distrusted by society, sousveillance is often directed *against* visible sources of surveillance (Mann and Ferenbok, 2013). By recording surveillance, people resist and draw attention to the surveillant assemblages of the state. Officials and law enforcement, in turn, hide their surveillance efforts—the ability to record in secret, with impunity, is often a necessary component of successful surveillance. And the dragnet data collection of surveillance companies like Facebook and Google has enhanced capabilities for both surveillance *and* sousveillance (Zuboff, 2019).

Conversation is also part of informal surveillant assemblages. With gossip and whisper networks, people can spread knowledge about the powerful outside of their grasp, as well as reinforce group bonds and boundaries (Lagalisse, 2013; Gluckman, 1963). These conversations become part of a backstage script, a 'hidden transcript', that only people within a community can access (Scott, 1990). Without being publicly uttered, those hidden transcripts then change how public conversation unfolds.

These forms of peer monitoring, or lateral surveillance, do have benefits (Andrejevic, 2005). The architect Jane Jacobs (1992) supported urban design that enabled informal community surveillance. She supported elements like short street lengths to increase the visibility of intersections and attractions that would draw people to a neighborhood at different times of day. Jacobs felt that a close-knit community watching out for each other was safer than formalized state surveillance. But lateral surveillance harms as well. People direct surveillance toward what they feel is most suspicious. Groups like neighborhood watches can disproportionately surveil people seen as outsiders, which often reinforces social prejudices (Lub, 2018).

And because people know their personal reputation is at the mercy of others, a culture of lateral surveillance encourages careful reputation management (Solove, 2008; Andrejevic, 2005). Sharing the wrong information can expel someone from a community, rather than bringing them closer to it. Modern states sometimes promote suspicious lateral surveillance as the social duty of responsible citizens (Reeves, 2012). In this way, communal surveillance leads to the same kind of impulsive self-control as Bentham's panopticon, a theme that will reappear in the section on digital vigilantism below.

2.3 Internet Communities and the Enforcement of Norms

The Internet's precursors, like Usenet and bulletin-board systems, were primarily text-based. As a result, early accounts of computer-mediated communication assumed that the separation between speakers removed social cues from computer communication entirely (Spears and Martin Lea 1992). Users, however, quickly created ways of communicating social cues with tools like letter repetition,

intentional misspellings, and timing of message transmission, based on their existing knowledge of social cues (Kalman and Gergle, 2014; Herring, 2012).

Usenet communities were organized around interest groups and voluntary affiliations. But any participant on a board could act maliciously to arrest an entire community's discussion. They could clog message boards with endless, inflammatory arguments—the 'flame wars' that were the predecessors of griefing and trolling behaviors found online today (Jane, 2015). Information scientists have noted the complicated strategies that these communities developed to deal with unwanted participants (Herring *et al.*, 2002). System admins could kick users from servers at their discretion, and communities would often test newcomers. Trolling was one common tactic. Experienced users would post a piece of obviously incorrect information, and anyone who responded seriously had taken the bait and revealed their naivety by not understanding the group's in-joke (Tepper, 1997).

Susan Herring (1999) has sketched out several other strategies that users employed when confronted with people who acted against their community norms. (She developed her theory to describe women's responses to misogyny on IRC channels and email lists.) Herring noted that people targeted by harassment in a digital community tended to either escalate, accommodate, or fall silent. Harassers would force their targets to adopt one of these strategies, thereby reinforcing the dominance of their rhetoric. In cases of escalation, admins would usually step in and apply what was often arbitrary judgment because of how hard it was to interpret communal infractions.

With social media, similar problems exist, but the role of platforms in defining and moderating communities is much larger. Where before users had relied on system admins, they now rely on social media platforms' content moderation systems, which are applied both algorithmically and with human discretion. Aside from community agreements and the like, platforms also provide (some) tools for users to protect themselves against unwanted conversations, like block lists, muting subjects, and locking accounts down with privacy settings (Jhaver *et al.*, 2018). Where independent Internet communities are sometimes at a loss because they don't have an external authority to petition (Herring *et al.*, 2002), social media users can now make requests of platforms. The same problem of arbitrary mediation still exists, however.

2.4 Online Anonymity and Privacy

Anonymity and privacy are integral to online participation, but they have many unintended consequences. danah boyd (2012) in her proposal of 'networked privacy' argues that privacy, usually assumed to be an individual right, actually has group implications. In one example, she points out that when a person gives their DNA for sequencing, they generate information about their whole family as well. Privacy is also contextual—what we share with our doctors is not the same as what we share with employers or acquaintances (Nissenbaum, 2011). And results in computer science have shown how easy it is to deanonymize, or re-identify, anonymized data sets (Narayanan and

Shmatikov, 2009). All of these factors make protecting data privacy a complicated task. It requires more than just the 'informed consent' of users.

Internet communities, like the hacktivist collective Anonymous or the trolls found on sites like 8Chan, have complicated relationships to anonymity. For those groups, remaining anonymous is a key requirement for membership. On the one hand, anonymity prevents members from hoarding fame, ensuring some egalitarianism (Coleman, 2015: 190). On the other hand, it allows members to act without impunity towards other groups and one another. Anonymity grants trolls an asymmetrical relationship to their targets. Victims of trolling or mass harassment can't tie their attackers back to a concrete identity, just the mask of anonymity:

> *Trolls don't mean, or don't have to mean, the abusive things they say. They get to choose the extent to which their statements match their personal beliefs; they get to establish that they're just trolling* […] *Targets of trolling, on the other hand, are expected to take trolls at their word, and are only trolled harder if they resist.* (Phillips, 2015: 26)

As hinted at before, however, anonymity online is mostly an illusion. Users normally possess only a degree of pseudonymity. They operate under a username or nickname, but the data collected by websites combined with self-disclosures (like where they live, gender, or personal email addresses) often means it's possible to unmask the user behind a username. The problem of determining true anonymity has lead theorists to differentiate between technical anonymity, when a person can't be identified, and social anonymity, when a person or a community acts *as if* they can't be identified (Bancroft and Scott Reid, 2017).

Doxxing is a behavior that emerged from the disconnect between social and technical anonymity—the identification of a user's legal name, contact information, and physical address combined with the publication of that information on public websites like Pastebin or Reddit. Currently, the most complete theoretical account of doxxing is philosopher David Douglas' (2016) article 'Doxing: A conceptual analysis'. Based on a sociological description of anonymity by Gary Marx (1999), Douglas divides doxxing into three categories: deanonymization, based on identity information; targeting, based on location information; and delegitimization, based on credibility or character information. In each case, doxxing involves the collation of public information that removes 'a degree of anonymity' from someone with the intent of directing media attention towards them.

Douglas describes doxxing as having many possible motivations, from a desire to harm to a wish to expose wrongdoing. Researchers have debated doxxing's use in accountability processes, but it's also been a hallmark of cyberbullying and many large-scale online harassment campaigns (Buozis, 2017; Chen, Cheung, and Chan, 2019). Sometimes online communities use doxxing to police group boundaries as well (Dobusch and Schoeneborn, 2015). For example, since membership in Anonymous is defined by anonymity, identifying a user de facto revokes their membership.

In a digital environment that hinders anonymity, users have had to cope by protecting their identity and managing how their online personas relate to one another and their legal identity. Juggling all these identities is hard and requires technical skill to do well. Some people obfuscate or 'poison' their data trail with false results. Others carefully control their self-presentation on specific sites. Users can silo their different personas so they don't come in contact with one another or self-censor to craft an uninteresting (and un-incriminating) 'vanilla' persona (Pitcan, Marwick, and boyd, 2018). These attempts at resisting identification are mostly 'weapons of the weak'. Tools like identity obfuscation are 'haphazard and piecemeal, creating only a temporary window of liberty or a certain amount of reasonable doubt' (Brunton and Nissenbaum, 2011: 17). Groups like Anonymous, however, use these tools to develop complex ethical guidelines for supporting or harming one another (Colton, Holmes, and Walwema, 2017). On the other end, digital platforms have also moved towards policies of openness and transparency. Communities like Wikipedia operate under the assumption of self-disclosure, and social media companies like Facebook have encouraged the consolidation of online personas with user's legal identities. (For example, Facebook accounts using a 'fake name' are likely to be banned.)

All of these methods and the act of identification, or deanonymization, have ethical consequences that Internet users are still figuring out today. How much of our vulnerability online are we responsible for? How far should individuals go to protect privacy? Some people disapprove when others don't manage their identity adequately or feel distressed because of their own uncertain pseudonymity (Mikaela, Marwick, and boyd, 2018). Even amid calls for openness through Internet technology, some empirical results have suggested that people dislike indiscriminate identification, even if it has public interest (McNealy, 2017). As of right now, most users have to decide on their own how much to preserve the anonymity and privacy of others.

2.5 Digital Vigilantism

The Internet offered a new forum for organizing vigilantes and a new venue for their actions. In the vacuum of state authority, users have formed groups to combat cybercrime by creating false personas to track down online predators, proponents of fake suicide pacts, and other cybercriminals (Huey, Nhan, and Broll, 2012). Despite altruistic motivations, these groups often have difficulty collaborating with official law enforcement; they are seen as either illegitimate or ineffective. Some kinds of harmful online behavior, like cyberbullying or harassment, also don't receive much official police interest because they don't clearly fit into existing categories of crime (Broll and Huey, 2015).

One of the earliest cases of mass digital vigilantism occurred in 2006, when the Human-Flesh Search Engine formed in China to expose a woman who was video-recorded killing a kitten; her identity was revealed and she was fired from her job in less than six days (Nhan, Huey, and Broll, 2015). The Human-Flesh Search Engine, although organized online, often involved offline investigative efforts,

and users without much technical ability would contribute files and search results. Despite its checkered track record, the Search Engine became a concerted form of crowdsourced justice. Similarly, in Russia, the tradition of citizen courts transformed into video-recorded vigilantism. Russian vigilantes track people down for traffic violations, public intoxication, or unethical business practice; humiliate them; and then post recordings of that humiliation online—a 'spectacle of punishment' (Gabdulhakov, 2018).

In the United States, digital vigilantism is usually associated with the Boston Marathon bombing in 2013. After the bombing, a swarm of Reddit users pored through video and photo evidence from that day to try and identify the bomber (Nhan, Huey, and Broll, 2015). Users tried to create forensic reconstructions of the incident and to identify suspicious persons for the official police investigation. Still, the investigation lead to disastrous misidentification, as 'in several notable instances online discussion gave way to rampant speculation' (Nhan, Huey, and Broll, 2015: 353).

In his definition of vigilantism, Les Johnston (1996) describes it as a planned, voluntary action taken by private citizens that uses force or the threat of force to control crime or perceived social transgression. At the same time, Johnston also warns against an open-ended definition of vigilantism, because it depends on participants' cultural assumptions about crime and transgression, as well as their social status. Extending Johnston's work, sociologist Daniel Trottier (2017: 55) has defined digital vigilantism as 'a process where citizens are collectively offended by other citizen activity, and coordinate retaliation on mobile devices and social platforms.' People in groups like the Human-Flesh Search Engine 'weaponize visibility' to punish their targets by directing mass scrutiny towards them (reminiscent of the synopticon mentioned above). Trottier (2017: 63) does not see vigilantism as necessarily breaking from legal order; vigilantism is instead an example of citizens renegotiating their public roles and 'acting in a way they believe the state should'. Vigilantism can even be encouraged by law enforcement, when vigilantes are deputized to investigate and punish on behalf of the state (Walsh, 2014).

Trottier argues that digital vigilantism meets all of Johnston's criteria, but because of the novelty of the phenomenon and diversity of cases (and the cultures in which they appear), it's hard to say when the term can be applied correctly. Trottier sees digital vigilantism as a conservative force that recreates social prejudices. But it's not clear how instances of vigilantism that occur across contexts or in opposition to cultural norms fit his schema. For example, feminist activists have used public denunciations against misogynist trolls, particularly following GamerGate in 2015, and they acted more to protect themselves than to harm (Jane, 2016). In addition, the way that media picks up and amplifies a case of digital vigilantism can affects its impact much more than the initial act of sharing a denunciation, since the harm of digital vigilantism comes from its public visibility (Trottier, 2019: 7). The attribution of responsibility and guilt for digital vigilantism is hardly ever clear, for those who perform it or are targeted by it.

2.6 Conceptual Framework and Research Questions

After reviewing all these theories, I planned to use them to address a central research question through interviews of online investigators:

— What are the ethics of practice for dealing with the identifying information of other people? How do people decide what they will or won't publish about another person?

I thought that this complicated question about ethics and decision-making was too abstract on its own, so I created two smaller questions to help me build specific details and context. They are:

— What is the process of online investigation?
— How does online identification compare to information gathering or publishing methods like surveillance, doxxing, and traditional investigative journalism?

The theoretical categories covered in this section were not incorporated directly into my analysis, but I will revisit them in the Discussion section to compare them to my results. Instead, I decided to adopt a grounded theory approach to this study, whereby I would rely on the resulting data as my guide, rather than a theoretical framework.

3 METHODOLOGY

To answer the questions above, I decided to use qualitative interviews to uncover how investigators view their work, ethics, and effects. But, after my initial literature review, I didn't think I could use theory to design the interview study—I couldn't find theories that I felt precisely covered the area of research. Because I wanted to describe the resulting data rather than infer from them, I thought that an inductive analysis would work well. So, I employed the methodology of grounded theory to guide my work. With its attention to process and causality and its aim to 'create' theory rather than test hypotheses (Dey, 2004), I saw affinity between grounded theory and my research questions. I view my resulting study design as in line with Gabriella Coleman's (2010: 489) call for researchers to 'provincialize' culture in digital media: 'showing how, where, and why [digital media] matters is necessary to push against peculiarly narrow presumptions about the universality of digital experience.'

3.1 Sampling

My interviews started with a convenience sampling of respondents who were closest to my initial area of interest, doxxing and deanonymization. As interviews progressed, I used purposive sampling

to fill in gaps where I thought I needed more material. For example, I made sure that my later interviews included archivists and activists, because I was curious how their outlook would reflect in their practice, in comparison to the journalists that I spoke with earliest. This practice is consistent with 'theoretical sampling' in grounded theory, where researchers continually refine their sampling choices as they collect data (Charmaz, 2006: 96).

Ultimately, this sampling approach meant that the aim of my study would be the production of theory and description of the data I had collected, rather than results that would extend to a larger population. This limitation matches Charmaz's (2006: 101) methodology of grounded theory, where theoretical sampling 'pertains only to conceptual and theoretical development; it is not about representing a population or increasing the statistical generalizability of your results.' I actually thought that this would be a useful bound for my study, since I planned to interview participants from several countries, as well as differing professional and personal backgrounds. Figuring out what population they represented would have been a difficult interpretative task in itself. The small sample size of my study (seven respondents) also limits its scope, so my desire was to provide examples of diverse behaviors and beliefs and make specific comparisons instead.

3.2 Interviews

I still had to decide how respondents would help me answer my research questions. Knowledge about culture and process is often what Michael Polanyi (2009) described as 'tacit knowledge'—knowledge that we know, but cannot tell, because it is implicit in our actions. To help me uncover tacit knowledge, my topic guide was inspired by James Spradley's (1976) ethnographic interview advice, as well as a preliminary discourse analysis of deanonymization events in news media. Spradley encourages interviewers to ask 'grand tour' questions that go through all the steps of a process and to inquire about respondents' semantics and word usage. I thought Spradley's approach would help me keep my line of questioning open to new topics, while still allowing me to contrast respondents' answers to one another. Even though I would not be conducting an ethnography, I felt justified in using an ethnographic approach to interviewing because ethnography and qualitative interviewing have often influenced one another and drawn from the same literature, as Warren (2002: 85–86) points out.

One of the limitations of interviews is that researchers have to beware the 'attitudinal fallacy'—respondents might unknowingly describe events or causal relationships incorrectly (Jerolmack and Khan, 2014). Interviews are best for relating the emotional responses and opinions of the respondents, which fit the scope of my central research question.

During interviews, I followed Spradley's advice to ask descriptive questions frequently and restate respondents' answers; both practices helped me make sure that I had understood the respondents' explanations correctly. I included a set of questions for contrasting terms and concepts in my

interview guide, to help me tease out differences in respondents' semantic meaning (see Appendix 1). I found these helpful techniques for encouraging respondents to address my own biased assumptions.

I conducted interviews in-person and over video-call software, and I used a small digital recorder to minimize the number of locations where audio data was stored. Where possible, I preferred to arrange in-person interviews, because they made it easier for respondents to include demonstrations in our conversation. For example, one respondent used their phone to show me how password-reset abuse can uncover the personal email addresses of social media users. During interviews, I kept my topic guide for reference, but I did not adhere to the guide too closely. I wanted to avoid unintentionally leading respondents to certain answers or themes. This did, however, make interviewing more difficult—in cases where respondents took fifteen to twenty minutes to answer a given question, it was hard to keep track of what topics had already been covered in the discussion.

3.3 Analysis

Because I started from interest in a process, I thought it best to use methods that would help me accurately describe that process before I categorized it. I planned to apply thematic analysis, modified by grounded theory, to the interview transcripts. I had conducted an initial literature review before analysis, but because I employed open coding, none of the themes or theories from the literature were used to create a coding schema. Emerson, Fretz, and Shaw (2011: 175) describe open coding:

> [Q]*ualitative coding does not start from pre-established or fixed analytic categories but, rather, proceeds inductively by creating analytic categories that reflect the significance of events and experiences to those in the setting.*

So, I started analysis by going through transcripts of the interviews and marking out themes as they appeared in phrases or events, line-by-line, as the respondents described them. That way, I stuck as close to the data as possible and only coded themes that emerged from the data, rather than comparing transcripts to a set of theoretical categories (Charmatz, 2006: 72–73).

For the initial pass of open coding, I read through interviews one at a time and annotated them with codes as I noticed them (see Appendix 2 for examples). I listened to interview audio as I went to make sure my codes reflected the intonation and implications of the speakers. After each interview transcript, I made short coding memos by attempting to directly answer my research questions from the perspective of the interviewee (Emerson, Fretz, and Shaw, 2011: 185–186). Coding memos were particularly useful since I kept the interviews open; they helped me to synthesize information that was dispersed across an entire transcript. I later used those memos to compare themes across participants and to figure out where the similarities and differences lay in their responses. I also read

interviewee's published outputs from investigations, especially those they mentioned in the interviews, so that I could compare them to the testimony they had given.

For the following rounds of analysis, I applied selective coding by picking out larger themes and occasionally connecting them together as categories or subcategories of one another. I followed Charmaz's (2006: 60–63) reticence for axial coding, however. I did not fully map out how these categories related to one another, because I thought the themes too intertwined for axial coding to produce meaningful results.

3.4 Ethics and Reflexivity

At a minimum, the ethical design of a study should prevent participants, including the researcher, from experiencing harms or loss of privacy from the research process (Bertrand and Hughes, 2018: 22). Early on, this tenet lead me to change my area of research. At first, I wanted to interview people affected by online harassment, particularly people who had been doxxed and their personal information published online. Interviews, however, can be emotionally distressing experiences for respondents (Warren, 2002: 89). I was worried about re-traumatizing people by asking them to relive their experiences during in-depth interviews or *through* contacting them out of the blue.

I thought about arranging focus groups to make the study more useful for those affected by online harassment, but the time and difficulty of the logistics proved prohibitive. Instead, I chose to remove my concern entirely. I decided to speak to people whose work involved acquiring, organizing, and displaying the personal information of others. I felt that talking to them about their methods and motivations would get me closer to my initial research interests and would help me address a gap in the literature.

At first, because I thought that my research might include participants from Internet subcultures like trolling, I was worried about risks to my own privacy. I undertook the removal of identifying information about me from people search and public record sites, as well as search engines. This process helped during interviewing because, out of curiosity, one respondent did indeed try to find my social media accounts, which added to our discussion.

To be sure I had respondents' informed consent, I discussed the study with them and provided an information sheet about the study design, with my personal motivations and sample questions included. I received verbal or written consent before each interview and offered to provide a copy of the transcript afterwards, so that participants could remove details if they didn't want them included in the study. For data storage, I numbered the audio and transcript files and used a separate password-locked file to correlate file numbers with the interview participants. I made sure to keep the audio files and transcripts stored in an encrypted drive.

I also planned to anonymize this study to reduce any potential harms. To minimize the collection of unnecessary personal information, I avoided asking for personal details that didn't relate to the study. That way I reduced the information I would have to redact from the transcripts. This is odd from the perspective of traditional interviewing practice, but researchers of Internet communities like Gabriella Coleman (2015) and Whitney Phillips (2015) have argued that avoiding structured questions about demographics or identity doesn't hinder qualitative analysis and in many cases is necessary for online research.

I anonymized the resulting transcripts by removing names of people, locations, places of employment, or specific media publications in ways that preserved the content but not the references of our conversation. Because of their media visibility, however, several respondents expressed scepticism about the efficacy of anonymizing their information (they would argue 'pseudonymizing') or about the need for anonymity at all. Furthermore, the anonymization of interviews depends on the judgment of the researcher and their ability to notice identifying associations, as Shklovski and Vertesi (2013) point out. I have included more reflections on this in the Discussion section below.

Finally, I have deliberately written portions of this paper in ways some might consider personal or informal. By adopting the first person, I could clarify my methodological and theoretical commitments and use active voice for verbs, which reduces sentence complexity. I have also avoided nominalizations where I did not need them, and I preferred to use vocabulary that would be more familiar to readers. These are standard ways to make research more accessible in technical communication and social science (Billig, 2013; Spyridakis and Wenger, 1992).

The design and ethics of this study were approved by my dissertation advisor in the Department of Media and Communications and by the Research Ethics Committee at the LSE.

4 RESULTS

The interviews ultimately comprised 10 hours of audio split over seven respondents. Respondents lived or worked in the United States, the United Kingdom, or Singapore, and most were journalists, hackers, activists, researchers or advocates (several in some combination). All respondents had practiced online investigations for at least several years, including investigators who identified people outside of work. Because of the duration of their practice, all investigators had adjusted their approaches to investigation and publication since they started. I'll discuss their answers to my research questions here and how they relate to the literature in a later section.

As mentioned in the methodology section, this paper did not involve participant observation. Therefore, conclusions about the actual actions of the investigators have to be limited. This came up

in one interview where a respondent claimed that a published report was a good representation of their process, but when I asked them about the steps of the investigation, they responded,

> *That's how it's laid out, chronologically on there, but it wasn't—I didn't, the actual process of finding things was definitely not in the same order as what you see on there.*

Interviewees mentioned how their insights often occurred simultaneously, making them hard to document. Ideally, future research into digital investigation would use participant observation to supplement the respondents' perspectives.

4.1 The Process of Investigation

Investigators were often prompted by a feeling of either curiosity or suspicion. Both could be sparked by learning about a person or event directly or through media. Several worked with groups that supported journalists, activists, or victims of online harassment and would undertake an investigation at their request. Investigators mentioned scrutinizing media events that they thought had dubious authenticity; others were motivated by a desire to preserve information. Investigators who acted as part of their occupation didn't cite curiosity as a motivator as often. For those who mentioned curiosity, the challenge of identification was also a motivation. As one respondent described,

> *It's like figuring out, it's like cracking a safe, but not stealing any money. For me, it's the cracking of the safe that's fun. I don't want the money inside.*

Other investigators also mentioned being motivated by a sense of justice—a desire that the information they obtained would lead to societal improvement or the reparation of a specific injustice. A respondent who researched police misconduct explained by saying,

> [I]*t causes real issues with people. Trauma, basically. So people… people want the truth.* […] *you've been betrayed by someone but that you've in fact been betrayed by the state and that it was not just a person that betrayed you, but there was an entire hierarchy of people behind it* […] *I think getting to know the truth is both important for dealing with it at a personal level and also there, there is a quest for acknowledgment and transparency and sort of, apology is just the first step.*

In cases where investigators were tracing systemic injustice, the identification of a person was consequential only to the extent that it revealed a social pattern—the individual wasn't as important as the social network.

Interview respondents described using a combination of the following techniques to collect information or draw conclusions:

— matching collected data to social media accounts;
— tailoring search engine queries (a method sometimes called 'Google dorking');
— visually identifying people in photographs;
— comparing background details in photos to satellite imagery or maps to geolocate the source of a photo;
— hacking target's email accounts or websites;
— making public records requests;
— reverse-image searching on sites like Google and Yandex to find either the origin of a photo (which could indicate its authenticity) or where an image had spread online;
— web scraping to preserve sites or massive datasets;
— crowdsourcing investigations by discussing leads on social media sites;
— using approximate results like an app's 'users near you' function to narrow down guesses;
— learning what institutional records would exist within a target's 'document cloud' to tailor future searches (Cuillier and Davis, 2019); and
— playing with a site's functionality to see if identifying information, like email addresses, is hidden in its user interface.

At the outset, respondents followed different steps based on the tools and information accessible to them. Several described a process of 'taking inventory' before searching for new information. Investigators who had access to automated storage tools tended to pull an entire dataset for perusal and then look for connections afterwards. Mass downloads also let investigators preserve copies for their reference, in case a site went down. Investigators who could only acquire data piece by piece adapted their search based on each new bit of information. Respondents might either go through all the results of a search engine or web scrape to look for links, or they might find a piece of unique information, like a name or address, and work backwards, looking for more information to match it to.

Most interview respondents commented on the ease of acquiring identifying information about their subject: it didn't take long and it didn't require much technical knowledge. Several investigators balked at the idea of even calling the sites and data repositories they use 'tools' at all. (Future research might check how accessible these methods actually are for most Internet users.) As one interviewee explained,

> *Yeah. So this isn't like some superduper spy—this took like four hours, it wasn't like some extremely long, in-depth investigation. It was just running relatively simple searches and following the threads and throwing it together in one place.*

In some cases, demonstrating that ease was the investigator's goal—they wanted to warn specific people that they were identifiable online or to show police and media this structural problem with the Internet.

Investigations also often stumbled on connections that identified people other than their target; that reached a scope larger than the investigator originally realized; or that contained disturbing or sensitive information. This usually happened when an investigator was hoping to attribute responsibility for an event, not just identify a particular person. One respondent described their surprise at the change in scale of their investigation:

> *I basically ended up enumerating the anatomy of a large-scale WordPress botnet attack from scratch over an afternoon. And it was just like, 'This is, would be very easy to do.'*

Another investigator, who was attributing criminal acts, realized that their subject now had a political appointment.

A minor theme in the interviews was that respondents valued verbal confirmation above the data sources they found online. To encourage a subject to confirm their suspicions, respondents might either publish some of their results to prompt some kind of confession or subtly direct conversations towards an undisclosed topic that they'd researched. Conversely, verbal testimony could be a damning contradiction for an investigation. Along those lines, one respondent expressed a preference for working only with historical records to avoid writing about people who are still living.

To conclude investigations, respondents had several outlets. For most, particularly the journalists, their aim was to write their results into a media publication. Others, however, collaborated with journalists to publish news pieces as a way to pressure governments and other groups into addressing the issue they had uncovered.

The presentation of this information also involved making ethical decisions: I'll address both topics more in the Ethics of Practice and Publishing section below. Those looking to repair individual wrongs, however, didn't seek media outlets. Similarly, those motivated by curiosity often didn't do anything with the results, although the unintended connections made during investigations sometimes lead them to report someone's identifiability or potential harms. Those connections created moral obligations for the investigators—to act when they otherwise wouldn't have.

Some investigators created databases or resources so others could repeat their process or learn from their results. For example, one investigator created a questionnaire template to track information that justified or disproved their initial suspicions. Another was building a database of collected information so that they could carefully parcel out incidents and refer them to local media or law enforcement.

Many respondents expressed concern about the perceived legitimacy of their methods. To explain their approach, they often drew on comparable traditions, like corruption reporting in Russia, 'funa' denunciation in Chile, and the policies of Wikipedia editing. Some investigators published 'open source' media pieces that documented the process of investigation and all the sources they were drawing from. That way they could encourage others to use those methods, as well as demonstrate their validity. As one journalist describes,

> [T]*his methodology and way of presenting, researching, and collecting, verifying and presenting information through these, you know, totally digital means gets more public awareness, and gives a little bit more legitimacy, because a lot of people try to wave it away, like, you know, 'Eh, you know, these are just Twitter posts,' right?* […] *which yes,* […] *but you know, you can still get a lot of information from it, and when places with the production values and profiles of the BBC and* The New York Times *publish this stuff the water rises and everyone comes up.*

This rhetoric was also followed by an investigator who worked primarily with public records and Freedom of Information Act requests, although they felt that they had limited success in encouraging readers or other journalists to adopt their methods. Respondents generally avoided making any tentative results or partial investigations publicly available out of fear that they would be misinterpreted or that the validity of their methodology would be scrutinized.

4.2 Perception and Comparisons

Most of the investigators rejected comparisons to digital vigilantism and websites that dump data online indiscriminately, like Wikileaks. In addition, none of the investigators saw their work as comparable to doxxing. For one journalist, they would only say they had doxxed someone if they had made a mistake and accidentally identified a source in a published piece—doxxing was never the intention of their reporting. For others, they did not regularly communicate their findings. In that case, a sense of personal responsibility differentiated them:

> *When I find information on someone, if I do use it, I act on it myself. A lot of time with doxxing, it feels almost like the person wouldn't dare do anything on their own and they're putting this information so that everyone… so that an angry group or even a mob can then attack the person or hurt the person.*

One of the respondents mentioned the crowdsourced investigations of the Boston Marathon bombing as an example of a misuse of digital investigations. Respondents set themselves apart from these cases because they saw themselves as providing detailed, contextualized information, not just isolated, verified facts.

Several respondents differentiated themselves from mass surveillance because of the targeted nature of their investigations—their desire to investigate preceded information gathering. This view related to the reticence in publishing preliminary results mentioned above. As one respondent put it, while explaining why they disliked the term 'naming and shaming',

> [T]*he entire base of our work is to, to fight against rumors and unfounded accusations. Um, so that is the last thing we want to do online.*

For the most part, respondents saw their investigations as an extension of their job. So, they tended to compare themselves to other people in their field, like journalism or academia, even if their methodology set them apart. Several made comparisons to journalists in cybersecurity or archival researchers, even if they weren't part of those communities. One respondent in particular described investigation as a standard part of social acquaintance—that getting to know someone involved learning more about them on social media and search engines.

4.3 Ethics of Practice and Publishing

All respondents described having ethical quandaries during investigation or publication. The problems they mentioned most were those of collaboration, online harms, and the context of publications, and investigators developed tactics for enacting their ethics in all of those domains.

As mentioned before, investigators worried about the perceived legitimacy of their methods and had problems with people, particularly law enforcement, not understanding or believing their approach. In cases of lateral surveillance or sousveillance where investigators looked into peers or authority figures, investigators had to be even more careful about making conclusions so their subject wouldn't invalidate their results. To compensate, respondents collaborated to spread their results and relied on parallel investigations as an additional layer of verification. But to avoid sowing speculation, investigators needed to trust their collaborators before they could share preliminary results or documentation. Collaborators who handled data ethics differently than a respondent, however, posed a problem, since those differing opinions weren't always immediately obvious.

For example, activist respondents cited the difficulty of working with journalists—the journalists could only publish a small subset of the results; they often had to repeat information from piece to piece; and the constant pressure to publish meant that most journalists could not contribute much to long-term investigations. One respondent had their personal identity inadvertently published by a journalist collaborator along with information about an investigation, and they suffered severe repercussions as a result. Other investigators talked about the difficulty they had with crowdsourcing investigations: they had to limit crowdsourcing on sensitive investigations because they worried about the harm that media visibility of the investigation would cause.

Parallel state or law enforcement investigations were also mentioned several times. The closed nature of those investigations and the suspicion they had for investigators' independent research, however, meant that respondents had a hard time collaborating with them (a result in line with Huey, Nhan, and Broll, 2013). Instead, law enforcement investigations mostly served as external evidence to vet results, whenever they published conclusions. Still, some respondents had luck sharing results with law enforcement, although they were often unsure of the effects of their disclosure. This came up in cases where respondents felt an ethical obligation that *something* concrete be done with their results, like when they had uncovered instances of crime. One respondent described the responsibility they felt by saying,

> *It's not enough to put the information out there, even in the hands of the police and media. There needs to be also given additional resources—'This is what happened, this is what you should do. This is what some people have done historically. These are some of the different consequences. Also, please get in touch saying that you're OK, so that I can tick you off the list as well.'* […] *I mean, maybe I could have just dumped it on the Internet. It was not effective at even getting meaningful interviews, dumped. It didn't work. Maybe it would, if I did it today, it would? But I'm going to do it properly, and so when it's done, it's done. That's the idea. It'll be completely wrapped up, and I'll have a list of names. I'll be ticking them off. Tick, tick, tick, tick. 'Have police reference, Person got in touch. Person says thank you. Person says fuck off.' Tick, tick, tick, tick.* […] *until it's done. For me, it's the only way, responsible way of… doing it.*

The asymmetrical transparency of collaboration (where respondents disclosed their information, but information wasn't disclosed to them) often made them doubt whether just sharing information was enough to relieve them of their responsibility.

Differing technical skill also forced respondents to compromise when they shared results. One respondent had to switch to using a more insecure way of sharing information, namely a Google Docs file, when their collaborators refused to use an access-controlled database.

Respondents were all aware that their work could incite online harassment. All of them said they took measures to prevent harassment, even when their goal was to release information that would harm the reputation or delegitimize the subject of their investigation. Generally, they avoided publishing raw data, which is why publications or collaboration with media was such an integral part of most investigators' work.

To reduce the chance for harassment, investigators would redact information like names or addresses and blur photos. They paid especial attention to protecting the identities of people incidental to the investigation who appeared in photographs or texts they wanted to publish. One investigator described how they decided to redact a target's name because of the implications that publication could have for his family:

> [T]*his was* [a nationality] *surname of which there was only one family in* [country] *with that name. So if we would publish that, that would, that would interfere with his ex-wife, his then teenager kids, that would have, yeah, kind of repercussions for the entire family* […] *we had the confirmation* […] *yet we haven't released any of that until now, because we had the decency to keep that behind, to not do that.*

Across the board, the interview respondents thought the publication of identifying information was irresponsible, when it wasn't necessary.

Another technique that investigators used was explicitly hiding the sources of certain information or moderating the content they included in their databases. Obfuscating information about a source was important when revealing a source would endanger them—if they had published a video from a conflict zone, for example. In some cases, respondents would take down information at the request of people who were close to the investigated subject and upset by the publication (e.g., owners of archived sites or family members). That practice is in line with the emphasis that investigators had on the moral force of oral testimony:

> *This is, this is his own interp—his own saga of this experience* […] *we wanted to give him as much leeway to… push back against the narrative in the records as possible* […] *we're not publishing, you know, a book on this thing, so it's not like—we're not calling him up and making sure, verifying, making claims. We're just saying 'Here's what's in the file,' but that's only half the story.*

This kind of redaction or embargoing happened both when people contested the results of their investigations or expressed their desire to be removed from data collection.

These methods produced a further dilemma, however, for investigators who tried to publish open-source, or just well-contextualized, publications. Because, by necessity, all of their writing was supported by publicly available information, their audience could technically always find the source of a fact or photograph:

> [B]*y design it's easy to replicate our research. That's the whole point of it, right, is that people can replicate it and check our sources, and because of that, even if we take measures to blur faces and… you know, cover up names and all that, people can replicate our research and find these people and therefore go and harass them.* […] *we try to get them some measure of security and privacy, but also to the point where people know we're not fabricating and making stuff up out of thin air, if that makes sense. And there's no playbook with this.*

As explained above, how much protection an investigator should provide was a personal judgment tempered by the reduced verification of their results and how flimsy the source's anonymity was already. Sometimes investigators would also hide the source of their information when they thought

it would disturb their audience or harm their own reputation. All these protections, however, were antithetical to open-source publication (although not all investigators described their work as 'open source').

Investigators also considered the context of their work as part of ethical publication. They disliked information being taken out of the context they intended it to have. For example, another community could adopt the results of an investigation and use them as proof of a conspiracy theory. Or online users might use sharing features on social media that isolate and decontextualize posts as a way to generate online harassment:

> [Y]*ou start trying to like de—uncover dirt from people and you start retweeting that, and that's interesting because it's kind of—people are using that lack of context. Not only because you or me were less aware of the impact. It's also because we said things in a historic or a situational context, or while we were watching a movie and you just commented on the movie and then that thing, in itself, alone, ten years later, can of course look horrible.*

To prevent their investigations from being taken out of context, investigators would usually only release information to trusted collaborators. Several investigators also changed the way that they wrote about investigations; they would deliberately avoid making speculative statements or connections to sources of misinformation. This was difficult to do, especially with conspiracy theorists. Investigators could prove that records existed about something, but they often couldn't prove that records *didn't* exist about that same subject. One investigator stopped making jokes about investigations because those claims would be taken too literally. The dilemma of having little time to create the proper framing for investigations reappeared as well:

> *I wish I had more time and energy to sort of make sure there was enough of that context—I wish I could make people care about that context. I used to, you get, you're sort of stuck at the case where you're making people feel smart and, like, hip and with it by dangling this piece of information that makes them feel like they've read their books, or you can like, you know, punish them by making them actually do, like 'No, you have to actually continue reading, and you have to push through, and understand those concepts.'*

Respondents all experienced people using the results of their investigations to make dubious decisions. They expressed frustration at being at the mercy of readers who interpreted the investigators' work however they wanted or who came with drastically different assumptions and knowledge about privacy online. Occasionally they also felt protective of those who didn't know they were endangering their privacy online.

Investigators had a hard time balancing their moral obligation to release time-sensitive information with the lengthy work of trying to ensure that no harm would result from that publication. In many

cases, they viewed remaining silent as unethical; they felt that the knowledge they had about impending harms or the contribution they could make to government transparency compelled them to speak. But that doesn't mean they were certain about the choices they made—respondents stressed that their decisions about presentation and anonymization were case-specific. They might make different decisions, given the opportunity.

5 DISCUSSION

Going back to the literature review, this interview study has clear implications for theories of anonymity and online behavior, especially those that specifically deal with doxxing and digital vigilantism. The respondents' navigation of the tension between social and technical anonymity could contribute to ethical frameworks that classify online behavior and change how we view anonymity and privacy in online culture. Similarly, the examples provided by respondents and their reticence to use certain labels to describe themselves suggest that theories of doxxing and digital vigilantism have to be careful in how they make distinctions in cases like these.

5.1 The Social Practice of Anonymity

Anonymity is not antithetical to trusting communities; it isn't necessarily 'deceitful or harmful' (Bancroft and Scott Reid, 2017). As Marx (1999: 104–105) mentions, we often expect to be anonymous in public places and it's perfectly normal to be sceptical about unprompted requests for identification in those cases. Historically, many Internet subcultures formed around assumptions of social and technical anonymity, before mass surveillance was possible. Expectations of social anonymity have persisted in online communities, even when technical anonymity isn't possible—participants could find each other's identities if they wanted to.

As study respondents often experienced, intuiting another person's desire for privacy is hard to do, and that task is made even more difficult when that person has an unrealistic expectation of anonymity. You might think of yourself a private person, but you might forget to hang curtains on your windows. Are your neighbors obligated to treat you as if you did have curtains? Respondents often dealt with people who didn't understand how much information was available about them online. Those people didn't know how architectural elements of the Internet, like targeted advertisement, search-engine site crawlers, or internet archives, create auditable trails of online actions or how collections of metadata could reveal identifiable information, online or offline. In fact, it was often this misunderstanding that respondents took advantage of in their investigations—their targets were unaware that they could be tracked in that way.

When they found out, however, investigators sometimes had to respond to seemingly incoherent requests for privacy. One archivist respondent (who had to take down archived copies of personal web sites) viewed these requests as absurd even as they complied with them,

> [The people being archived] *think there's such a thing as a hybrid public-private space online* […] *And anything you do to violate that is a representation of your lack of humanity and misunderstanding their rules and I'm like 'Yeah, and all you're doing is completely misrepresenting the entire medium of the Internet, like… you want to shout from the rooftops, but you only want certain windows open for people to hear it. Fine?* […] *you're just postponing the inevitable. If we can do it on an industrial scale, anybody can do it on a personal scale.'*

In this case, the respondent thought their copies caused minimal harm to those who wrote the sites. Because this investigator was more visible than other sources of online tracking, people looked to them to resolve the issue. All the respondent could do, however, was restore their sense of social anonymity, not their actual, technical anonymity. This dilemma was mirrored in cases where respondents found surprising results in public records—like disturbing information in police reports or copies of conspiracy theory documents embedded in official records. They might have given disturbing material more public visibility, but anyone could have reached that material on their own.

The source of respondents' ethical concerns is telling—most worried about how they presented and stored information they had collected, but not about the act of collection itself. Sometimes that was because they hadn't intended to collect a certain kind of information. At other times, the information they were looking for had already been made widely available on the Internet. And in cases where they didn't plan to do anything with the personal information they uncovered, they didn't see the harm in finding out more. Because of the nonlinear path that investigations followed, respondents' ethical obligations arose out of the act of collection. In other cases, a separate event caused investigators to act when they might not have. In light of new circumstances, using information they had already collected became justified.

Both these phenomena—dealing with unrealistic estimations of anonymity and the ethical duties that follow learning certain kinds of information—could supplement theories of anonymity and privacy. For example, Colton, Holmes and Walwema (2017) have proposed an ethics of care framework for tactical communication in groups like Anonymous. Their work is based on instructional documents, and it would be improved by interview about the ethical motivations and concerns that inspired those documents. By using interviews about ambiguous expectations of anonymity, theorists will arrive at better descriptions of online cultures and better moral proscriptions for how those who procure information should act.

5.2 Doxxing and Digital Vigilantism

The results of this interview study pose problems for the literature on doxxing and digital vigilantism. As Johnston (1996) warned, definitions of vigilantism risk being too open. The phenomenon that this study describes—the seeking and publishing of personal identification online—is enabled by a culture of lateral surveillance and mass surveillance, but I do not think it is equivalent to either doxxing or digital vigilantism.

For doxxing in particular, theorists have started to use it outside of its origin in Internet subcultures. In frameworks like Douglas' (2016), doxxing describes a whole umbrella of behaviors that range from harassment to public-interest denunciation. This study suggests that people who perform most of the actions of doxxing in good faith still do not view their own actions as doxxing, however harmful they are for their target. It might still be a useful term, but having it as an organizing concept for a constellation of general behaviors would ignore its origin and the Internet cultures that most frequently use it. And for theoretical studies, it's worth noting that doxxing as a term is more likely to alienate those it's being used to describe—perhaps a more neutral term would serve overarching theories of deanonymization better.

This study raises similar concerns for digital vigilantism. While several respondents might qualify as vigilantes, some did not define themselves as seeking to cause harm. For others, their definition of justice did not involve punitive action—knowing and attributing responsibility was enough. Vigilantism also upholds societal norms, making it a fundamentally conservative force (Trottier, 2019: 2). People that enforce their own norms, separate from judicial standards, might not deserve the label of vigilante. The respondents in this study often defined their work in aspirational terms—they saw their work as contributing to a more just society. In some cases that might align with existing social mores, but in other cases the respondents were enacting a justice that they wish that their society recognized at all.

Sometimes, respondents demonstrated how duties of surveillance and documentation have been outsourced by law enforcement groups onto victims and other citizens. Building dossiers is no longer isolated to the state; people have to share that burden if they expect to receive services. As one participant who helps victims of online harassment described,

> [I]*f you want to get anything done in court or wherever, you need to keep the evidence, because many people, their instinctive reaction is to just go and delete everything and then you don't have any evidence of the material or things that could lead to identify people,* […] *It's more about documenting what they receive and systematize. First, not delete it, and second, systematize it in a way that would make sense for law enforcement.* […] *these things people keep can lead to identify someone, that's kind of the core advice, if they want the justice to jump in.*

With an overly inclusive definition of digital vigilantism, the wrong kind of behavior could be swept up and categorized as vigilante activity. It's true that the Internet has enabled more lateral surveillance and the use of media visibility to cause harm, but theories of digital vigilantism should account for the fact that vigilant watching is often executed under duress.

5.3 Anonymization in Research Design

One methodological result of this study has been, ironically, about the difficulty of ensuring participant anonymity. Several investigators with computer experience adopted personas or fake accounts and masked their activity reflexively, even if they didn't see anonymity as integral to their work. Because so many respondents had published media pieces (easily discovered through a search engine), it was hard to mention any details of their cases without violating their anonymity. Several respondents mentioned this, as part of their concern for the study's anonymization. And the preservation of participant privacy as set out in the ethical design of the study was challenged by participants several times as unnecessary. At other moments, my ethical duty became incredibly ambiguous. How should a researcher react when they have an elaborate data storage plan and then an interviewee reveals that they've had an Amazon Echo speaker running in the background during the interview, recording audio, and sending it to a cloud server?

Finn Brunton and Helen Nissenbaum (2011) have argued that researchers who promise total anonymity are often being unrealistic about the amount of privacy they can provide. Rather than completely preventing identification, researchers obfuscate participants. With the right auxiliary information, identification is still possible. Depending on context, the default assumption that social science data must be anonymized might be overzealous—both in the way it reduces the information a study can provide and how it unrealistically downplays the potential for participant identification (Tilley and Woodthorpe, 2011). For this study in particular, it was difficult to anonymize the writing so that it conveyed specific detail and did not make undue generalizations. Discussing the risks and harms of identification with each participant more and planning for an extended period of monitoring the circulation of a study might make for a more ethical and effective approach (van Baalen, 2018; Shklovski and Vertesi, 2013).

6 CONCLUSION

This study demonstrates the problems that arise when people collect and share personal information online. Investigators used a variety of means, including digital and physical sources, to identify people online, and they were motivated by curiosity, the requirements of their work, and their sense of justice. Their efforts to preserve anonymity were mostly confined to people peripheral to their investigation, although they also took care to prevent their work from causing online harassment.

Investigators used collaboration mostly for the publication of their results, although they had difficulties with the differing technical ability of their collaborators and audiences. As a result, they were protective of how their work was interpreted and tried to avoid publishing speculation to preserve their legitimacy and prevent online harms. The small sample size limits the scope of this study, but I hope it is an indicator of fruitful directions for future research in Internet ethnography, as well as a helpful example for theories of Internet culture.

I've suggested some useful limitations to theories of digital vigilantism, but future research could make connections to the fields of information retrieval and library science to enrich our understanding of how people answer questions of identification. The way respondents built dossiers and databases out of the information they collected could also provide a contemporary update to the paper-based media archaeology mentioned earlier. The 'open source' methodology of some respondents links to research on the politics of 'open' Internet communities like Wikipedia and the open-source software movement (Tkacz, 2015; Coleman, 2004). As some respondents demonstrated, openness can actually hide an act of omission. Finally, the limited cross-cultural comparisons possible in this study suggested that a broader examination of the ethics of online anonymity would help describe the already-legitimized practices that respondents used to contextualize their actions.

Anonymity underpins many social interactions, but it's easily revoked. Although I've tried to be neutral, the act of online identification is fraught, because of all the harms that it can cause. But it's also normal on the Internet, since so much identifying information is freely available to those with Internet access. The infrastructure of the Internet enables lateral surveillance and sousveillance, not just mass surveillance, and the investigators I interviewed all had to adapt their existing ethical standards to this new domain. Normative ethics in privacy policies and regulation have received a lot of attention recently, but I hope these interviews show that applied ethics on the Internet deserves more social science study. We need to talk more about how privacy and anonymity are performed socially—what their role is and how people decide to preserve or puncture them.

ACKNOWLEDGMENTS

First, I want to thank my parents Anthony DiBella and Marjorie Ball, and my partner, Alexandra Chipkin, for supporting me throughout this year at the LSE. They called when things weren't going well; they talked me through every twist and turn of this paper; and they believed in me. I can't thank them enough.

I also appreciated the help of my advisor, Dylan Mulvin, who debriefed me after each and every interview; Rimma Chipkin for her Russian translation; and the people at Privacy International and the Cypurr Collective for creating spaces to discuss the ideals and details of digital privacy. Finally, I want to thank all of the interview participants for being so willing to share their work and opinions with me. I hope that this representation does them justice.

REFERENCES

Andrejevic, Mark. (2005) The work of watching one another: Lateral surveillance, risk, and governance, *Surveillance & Society* 2 (4): 479–497.

Bancroft, Angus, and Peter Scott Reid. (2017) Challenging the techno-politics of anonymity: The case of cryptomarket users, *Information, Communication & Society* 20 (4): 497–512. https://doi.org/10.1080/1369118X.2016.1187643.

Beniger, James Ralph. (1997) *The Control Revolution: Technological and Economic Origins of the Information Society*, Cambridge: Harvard University Press.

Bertrand, Ina, and Peter Hughes. (2018) *Media Research Methods: Audiences, Institutions, Texts*, 2nd ed., London: Palgrave Macmillan.

Billig, Michael. (2013) *Learn to Write Badly: How to Succeed in the Social Sciences*, Cambridge University Press.

Broll, Ryan, and Laura Huey. (2015) 'Just being mean to somebody isn't a police matter': Police perspectives on policing cyberbullying', *Journal of School Violence* 14 (2): 155–176. https://doi.org/10.1080/15388220.2013.879367.

Brunton, Finn, and Helen Nissenbaum. (2011) Vernacular resistance to data collection and analysis: A political theory of obfuscation, *First Monday* 16 (5). https://firstmonday.org/ojs/index.php/fm/article/view/3493/2955.

Brunton, Finn. (2017) Notes from/dev/null, *Internet Histories* 1 (1–2): 138–145. https://doi.org/10.1080/24701475.2017.1307059.

boyd, danah. (2012) Networked privacy, *Surveillance & Society* 10 (3/4): 348–350.

Buozis, Michael. (2017) Doxing or deliberative democracy?: Evidence and digital affordances in the serial subreddit, *Convergence: The International Journal of Research into New Media Technologies*, August, 1–17. https://doi.org/10.1177/1354856517721809.

Chen, Mengtong, Anne Cheung, and Ko Chan. (2019) Doxing: What adolescents look for and their intentions, *International Journal of Environmental Research and Public Healt*h 16 (2): 218–232. https://doi.org/10.3390/ijerph16020218.

Charmaz, Kathy. (2006) *Constructing Grounded Theory*, London : Sage Publications.

Coleman, Gabriella. (2004) The political agnosticism of free and open source software and the inadvertent politics of contrast, *Anthropological Quarterly* 77 (3): 507–519. https://doi.org/10.1353/anq.2004.0035.

Coleman, Gabriella. (2010) Ethnographic approaches to digital media, *Annual Review of Anthropology* 39 (1): 487–505. https://doi.org/10.1146/annurev.anthro.012809.104945.

Coleman, Gabriella. (2015) *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*, London: Bloomsbury.

Coleman, Roy, and Michael McCahill. (2011) *Surveillance & Crime. Key Approaches to Criminology*, London: Sage Publications.

Colton, Jared S., Steve Holmes, and Josephine Walwema. (2017) From noobguides to #OpKKK: Ethics of Anonymous' tactical technical communication, *Technical Communication Quarterly* 26 (1): 59–75. https://doi.org/10.1080/10572252.2016.1257743.

Cuillier, David, and Charles N. Davis. (2019) *The Art of Access: Strategies for Acquiring Public Records*, 2nd ed., Los Angeles: CQ Press.

Deleuze, Gilles, and Félix Guattari. (1987) *A Thousand Plateaus: Capitalism and Schizophrenia*, Minneapolis: University of Minnesota Press.

Dey, Ian. (2004) Grounded Theory, pp. 81–94 in Clive Seale, Giampietro Gobo, Jaber Gubrium, and David Silverman (eds) *Qualitative Research Practice*, London: Sage Publications. https://doi.org/10.4135/9781848608191.d9.

Dobusch, Leonhard, and Dennis Schoeneborn. (2015) The communicative constitution of anonymous: Fluidity, identity, and organizationality, *Journal of Management Studies* 52 (8): 1005–1035. https://doi.org/10.1111/joms.12139.

Douglas, David M. (2016) Doxing: A conceptual analysis, *Ethics and Information Technology* 18 (3): 199–210. https://doi.org/10.1007/s10676-016-9406-0.

Duranti, Luciana. (1989) Diplomatics: New uses for an old science, *Archivaria* 28 (Summer): 7–27.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. (2011) *Writing Ethnographic Fieldnotes,* Chicago: The University of Chicago Press.

Foucault, Michel. (1995) *Discipline and Punish: The Birth of the Prison*, New York: Vintage Books.

Gabdulhakov, Rashid. (2018) Citizen-led justice in post-communist Russia: From comrades' courts to dotcomrade vigilantism, *Surveillance & Society* 16 (3): 314–331. https://doi.org/10.24908/ss.v16i3.6952.

Gangadharan, Seeta Peña. (2017) The downside of digital inclusion: Expectations and experiences of privacy and surveillance among marginal internet users, *New Media & Society* 19 (4): 597–615. https://doi.org/10.1177/1461444815614053.

Gitelman, Lisa. (2014) *Paper Knowledge: Toward a Media History of Documents*, Durham: Duke University Press.

Gluckman, Max. (1963) Gossip and Scandal, *Current Anthropology* 4 (3): 307–316. https://doi.org/10.1086/200378.

Guillory, John. (2004) The Memo and Modernity, *Critical Inquiry* 31 (Autumn): 108–132.

Haggerty, Richard D., Ericson, Kevin V. (2000) The surveillant assemblage, *British Journal of Sociology* 51 (4): 605–622. https://doi.org/10.1080/00071310020015280.

Haraway, Donna Jeanne. (1991) A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century.' In *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Herring, Susan C. (1999) The rhetorical dynamics of gender harassment on-line, *Information Societ*y 15 (3): 151–167. https://doi.org/10.1080/019722499128466.

Herring, Susan, Kirk Job-Sluder, Rebecca Scheckler, and Sasha Barab. (2002) Searching for safety online: Managing 'trolling' in a feminist forum, *The Information Society* 18 (5): 371–384. https://doi.org/10.1080/01972240290108186.

Herring, Susan C. (2012) Grammar and electronic communication, in C. Chapelle (ed.) *The Encyclopedia of Applied Linguistics*, Oxford: Blackwell Publishing. https://doi.org/10.1002/9781405198431.wbeal0466.

Huey, Laura, Johnny Nhan, and Ryan Broll. (2013) 'Uppity civilians' and 'cyber-vigilantes': The role of the general public in policing cyber-crime, *Criminology & Criminal Justice* 13 (1): 81–97. https://doi.org/10.1177/1748895812448086.

Jacobs, Jane. (1992) *The Death and Life of Great American Cities*, New York: Vintage Books.

Jane, Emma A. (2015) Flaming? What flaming?: The Pitfalls and potentials of researching online hostility, *Ethics and Information Technology* 17 (1): 65–87. https://doi.org/10.1007/s10676-015-9362-0.

Jane, Emma A. (2016) Online misogyny and feminist digilantism, *Continuum: Journal of Media & Cultural Studies* 30 (3): 284–297. https://doi.org/10.1080/10304312.2016.1166560.

Jerolmack, Colin, and Shamus Khan. (2014) Talk is cheap: Ethnography and the attitudinal fallacy, *Sociological Methods & Research* 43 (2): 178–209. https://doi.org/10.1177/0049124114523396.

Jhaver, Shagun, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. (2018) Online harassment and content moderation: The case of blocklists, *ACM Transactions on Computer-Human Interaction* 25 (2): 1–33. https://doi.org/10.1145/3185593.

Johnston, L. (1996) What is vigilantism?, *British Journal of Criminology* 36 (2): 220–236. https://doi.org/10.1093/oxfordjournals.bjc.a014083.

Kallinikos, Jannis. (2009) The making of ephemeria: On the shortening life spans of information, *The International Journal of Interdisciplinary Social Sciences: Annual Review* 4 (3): 227–236. https://doi.org/10.18848/1833-1882/CGP/v04i03/52870.

Kalman, Yoram M., and Darren Gergle. (2014) Letter repetitions in computer-mediated communication: A unique link between spoken and online language, *Computers in Human Behavior* 34 (May): 187–193. https://doi.org/10.1016/j.chb.2014.01.047.

Kittler, Friedrich A. (1990) *Discourse Networks 1800/1900*, Stanford University Press.

Komaromi, Ann. (2012) Samizdat and Soviet dissident publics, *Slavic Review* 71 (1): 70–90. https://doi.org/10.5612/slavicreview.71.1.0070.

Lagalisse, Erica. (2013) Gossip as Direct Action, pp. 112–168 in L. Phillips and S. Cole (eds) *Contesting Publics: Feminism, Activism, Ethnography*, London: Pluto Press.

Lenhart, Amanda, Michele Ybarra, Kathryn Zickuhr, and Myeshia Price-Feeney. (2016) Online harassment, digital abuse, and cyberstalking in America, *Data & Society*. https://datasociety.net/blog/2017/01/18/online-harassment-digital-abuse/.

Lub, Vasco. (2018) Neighbourhood watch: Mechanisms and moral implications, *The British Journal of Criminology* 58 (4): 906–924. https://doi.org/10.1093/bjc/azx058.

Mann, Steve, and Joseph Ferenbok. (2013) New media and the power politics of sousveillance in a surveillance-dominated world, *Surveillance & Society* 11 (1/2): 18–34. https://doi.org/10.24908/ss.v11i1/2.4456.

Marx, Gary T. (1999) What's in a name?: Some reflections on the sociology of anonymity, *The Information Society* 15 (2): 99–112. https://doi.org/10.1080/019722499128565.

Mathiesen, Thomas. (1997) The viewer society: Michel Foucault's 'panopticon' revisited, *Theoretical Criminology* 1 (2): 215–234. https://doi.org/10.1177/1362480697001002003.

McNealy, Jasmine. (2017) Readers react negatively to disclosure of poster's identity, *Newspaper Research Journal* 38 (3): 282–292. https://doi.org/10.1177/0739532917722977.

Narayanan, Arvind, and Vitaly Shmatikov. (2009) De-anonymizing social networks, pp. 173–187 in *2009 30th IEEE Symposium on Security and Privacy*, Oakland: IEEE. https://doi.org/10.1109/SP.2009.22.

Nhan, Johnny, Laura Huey, and Ryan Broll. (2015) Digilantism: An analysis of crowdsourcing and the Boston marathon bombings, *British Journal of Criminology* 57 (2): 341–361. https://doi.org/10.1093/bjc/azv118.

Nissenbaum, Helen. (2011) A contextual approach to privacy online, *Daedalus* 140 (4): 32–48. https://doi.org/10.1162/DAED_a_00113.

Phillips, Whitney. (2015) *This Is Why We Can't Have Nice Things: Mapping the Relationship Between Online Trolling and Mainstream Culture*, Cambridge: MIT Press.

Pitcan, Mikaela, Alice E Marwick, and danah boyd. (2018) Performing a vanilla self: Respectability politics, social class, and the digital world, *Journal of Computer-Mediated Communication* 23 (3): 163–79. https://doi.org/10.1093/jcmc/zmy008.

Polanyi, Michael. (2009) Tacit knowing, in *The Tacit Dimension*. University of Chicago Press.

Reeves, Joshua. (2012) If you see something, say something: Lateral surveillance and the uses of responsibility, *Surveillance & Society* 10 (3/4): 235–48. https://doi.org/10.24908/ss.v10i3/4.4209.

Reichel, Philip L. (1977) Dossier building as a social problem topic, *Teaching Sociology* 4 (3): 293–306. https://doi.org/10.2307/1316905.

Reigeluth, Tyler Butler. (2014) Why data is not enough: Digital traces as control of self and self-control, *Surveillance & Society* 12 (2): 243–54. https://doi.org/10.24908/ss.v12i2.4741.

Scott, James C. (1990) *Domination and the arts of resistance: Hidden transcripts*, New Haven: Yale University Press.

Shklovski, Irina, and Janet Vertesi. (2013) 'Un-Googling' publications: The ethics and problems of anonymization, pp. 2169–2178 in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, Paris: ACM Press. https://doi.org/10.1145/2468356.2468737.

Solove, Daniel J. (2008) *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet*, New Haven: Yale University Press.

Spears, Russell, and Martin Lea. (1992) Social influence and the influence of the 'social' in computer-mediated communication, pp. 30–65 in M. Lea (ed.) *Contexts of Computer-Mediated Communication*, London: Harvester-Wheatsheaf.

Spradley, James P. (1979) *The Ethnographic Interview*, New York: Holt, Rinehart and Winston.

Spyridakis, Jan H., and Michael J. Wenger. (1992) Writing for human performance: Relating reading research to document design, *Technical Communication* 39 (2): 202–215.

Steyerl, Hito. (2012) In defense of the poor image, pp. 31–45 in *The Wretched of the Screen*, Berlin: Sternberg Press.

Tepper, Michele. (1997) Usenet communities and the cultural politics of information, in D. Porter (ed.) *Internet Culture*, Routledge.

Tkacz, Nathaniel. (2015) *Wikipedia and the Politics of Openness*, Chicago: University of Chicago Press.

Trottier, Daniel. (2017) Digital vigilantism as weaponisation of visibility, *Philosophy & Technology* 30 (1): 55–72. https://doi.org/10.1007/s13347-016-0216-4.

Trottier, Daniel. (2019) Denunciation and doxing: Towards a conceptual model of digital vigilantism, *Global Crime*, March, 1–17. https://doi.org/10.1080/17440572.2019.1591952.

Tilley, Liz, and Kate Woodthorpe. (2011) Is it the end for anonymity as we know it?: A critical examination of the ethical principle of anonymity in the context of 21st century demands on the qualitative researcher, *Qualitative Research* 11 (2): 197–212. https://doi.org/10.1177/1468794110394073.

van Baalen, Sebastian. (2018) 'Google wants to know your location': The ethical challenges of fieldwork in the digital age, *Research Ethics* 14 (4): 1–17. https://doi.org/10.1177/1747016117750312.

Vismann, Cornelia. (2008) *Files: Law and Media Technology*, Stanford University Press.

Walsh, James P. (2014) Watchful citizens: Immigration control, surveillance and societal participation, *Social & Legal Studies* 23 (2): 237–259. https://doi.org/10.1177/0964663913519286.

Warren, Carol A. B. (2002) Qualitative interviewing, pp. 83–101 in Gubrium, Jaber (ed.) *Handbook of Interviewing Research: Context and Method*, edited by Jaber F. Gubrium, London: Sage Publications.

Zuboff, Shoshana. (2019) *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*, London: Profile Books.

APPENDIX 1: INTERVIEW TOPIC GUIDE

History of Practice

— Could you describe the organization you work with and the role you have within it?
— Was there anything that first attracted you to exposing, unmasking, or documenting certain people online?
— What do you think about the way that it's been covered in the media? Are there any particular stories that have stood out to you?
— Overall, what are your motivations?
— How well do you know the people that you write about? Are they people that you have met offline?
— What effect did you hope it would have? On the person? On their community?
— Who do you think uses or reads your dossier articles? Who is your intended audience?

Process Description

— Could you describe the process of how you collect information on someone online?
— What kinds of information would you include in the articles? What information would you leave out? How did you collect and store it?
— How did you prepare the articles for publication? And where did you end up posting or publishing them?
— Do you work together or alone? How would you discuss the articles with others?
— What did you do to make sure the files were correct?
— Altogether, how difficult is it? Does the difficulty vary a lot from person to person? What makes someone difficult to deanonymize/research?
— What limits did you choose for what you would put into the posts or where/how you would post the information?
— What precautions did you take for yourself or for the person you were writing about?
— What were usually the actual consequences of publishing or researching? Are there any examples that stand out or that were particularly unusual?
— Did you have any further contact with people you write about? If so, what were those interactions like?

Ethics

— What role do you think anonymity plays on the Internet? Or what role should it play?
— Right now, deanonymization is in kind of a legal grey area. How do you think we should treat or view it?
— How would you define/describe your practice? Is 'doxing' a fair word to describe it? What would be a 'typical' case?
— What do you think about other people who deanonymize others online? Are there any people whose work you particularly admire or disapprove of?
— While doing my research, I've found a lot of news articles and blog posts by people who claim that we should never deanonymize people online. If someone were to challenge you on the ethics of that, how would you respond? What would you say to them?
— Do you have anything else you would like to add, or any questions for me?

Descriptive Questions

— Grand Tour (take me through it)
— Example (give an example of)
— Experience (any interesting experiences)
— Native-Language (how would you use a term? how would you refer to this? what if, how would?)

Structural Questions

These should be repeated, to provide context:

— What are all the different kinds of X…?
— Can you think of any other X?
— Is it true that X is a Y? (verification)

Contrast Questions

— Is it true that X is a Y, but W is a Z? (verification)
— What are the differences between an X and a Y?
— Of X, Y, Z, which are most alike and which are most different?
— Ask interviewee to order lists of traits/words/etc

General

- Avoid closed questions
- Avoid leading in wording of questions
- Define terms
- Don’t indicate judgment of interviewee opinion
- Ask follow-up questions. Restate answers and verify. Invite more, rather than challenge

APPENDIX 2: THEMATIC ANALYSIS

Note: The table below lists the themes that were identified after an open-coding process. The codes were first grouped together and then labeled. An extended grounded theory project would take those themes and expand them into an abstract theory of the subject matter. Because of the generative nature of open coding, not all codes have been included.

Table 1: Themes and Codes

Themes	Codes
Collaboration and Context	Difficulties of collaboration with different groups, publishing platforms
	Need for visible verification in publications
	Misinformation by lack of context
	Same information used by different parties for different aims
	Differing professional standards/ethics
	Dealing across groups when they all structure their information and access differently
	Frustration with being taken out of context
	Difficulty of recovering context of historical documents and language
	Collective vs. individual or mass action
	Misinterpretation of evidence
	Misinterpretation or misappropriation of technical terminology (like 'cyber-')
	Human negotiation in earlier forms of mediated communication
	Netiquette and community education

Social Practice of Privacy and Identity Online	Relationship between administrative control and community norms
	Process and control as power
	Embargoing ('canceling') as a community practice
	People not talking about privacy; privacy as a tacit practice.
	Privacy as theatre, 'Folk knowledge'
	Responses to changing Internet culture
	Privacy-preserving behavior as counterintuitive to the lay user
	Awkwardness of bringing personal information into conversation
	Dummy accounts, 'records-squatting'
	Obfuscation methods as a link
	Privacy as knowledge of practice
Ethics and Investigative Practice	Difficulty of keeping an investigative piece open source
	Frustration at not knowing effects and results of investigation
	Using obfuscation to protect sources
	'Technically' available information
	Assumption that media coverage leads to public understanding
	Desire to seek the truth
	Trustbuilding vs. betrayal of trust
	Proof by contradiction

	Encouraging confession
	Staying anonymous during investigation
	Curiosity
	Right to store data & copyright, intellectual property
	Using investigation for personal gain
	Ethics of data dumping
Files & Sources of Information	Public records as infrastructure
	Creating references for others
	Creating an audit of investigation
	Physical locations or files for information
	Importance of verbal testimony
	Files embedded in files
	Rumors, gossip, and suspicions as start to investigation
	Process of external verification
	Comparing redacted versions of same file