

Excavating awareness and power in data science: A manifesto for trustworthy pervasive data research

Big Data & Society
July–December: 1–12
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517211040759
journals.sagepub.com/home/bds



Katie Shilton¹ , Emanuel Moss^{2,3}, Sarah A. Gilbert¹,
Matthew J. Bietz⁴ , Casey Fiesler⁵, Jacob Metcalf³,
Jessica Vitak¹ and Michael Zimmer⁶

Abstract

Frequent public uproar over forms of data science that rely on information about people demonstrates the challenges of defining and demonstrating *trustworthy* digital data research practices. This paper reviews problems of trustworthiness in what we term pervasive data research: scholarship that relies on the rich information generated about people through digital interaction. We highlight the entwined problems of participant unawareness of such research and the relationship of pervasive data research to corporate datafication and surveillance. We suggest a way forward by drawing from the history of a different methodological approach in which researchers have struggled with trustworthy practice: ethnography. To grapple with the colonial legacy of their methods, ethnographers have developed analytic lenses and researcher practices that foreground relations of awareness and power. These lenses are inspiring but also challenging for pervasive data research, given the flattening of contexts inherent in digital data collection. We propose ways that pervasive data researchers can incorporate reflection on awareness and power within their research to support the development of trustworthy data science.

Keywords

Big data, data science, computational social science, social media data, personal sensing, digital phenotyping, research ethics, research methods

1. Introduction

In the era of ubiquitous digital devices, researchers are increasingly able to draw conclusions about people's health, habits, beliefs, and practices using methods that require no contact with, or awareness by, research subjects. Increasing *datafication* (Cukier and Mayer-Schoenberger, 2013) renders ever more domains of human activity as digital data for the purpose of making useful inferences and valuable predictions about that activity. *Pervasive data research* benefits from the result: rich information about people generated through digital interaction and available for computational analysis. Examples of pervasive data research include social media research, passive sensing research, personal sensing research, digital phenotyping, and computational social science using search histories, geolocation data, or wearables that record and/or sense personal behavior.

Pervasive data researchers in academia, industry, and government face a set of nested ethical problems that emerge

from the combination of the datafication of human activity, growing mistrust of digital research practices, and mismatched norms between datafication realities and the traditional importance of research participant autonomy. From the perspective of research ethics, the notable change is not the “bigness” of digital datasets, but the ubiquitous nature of the data sources and collection methods that allow researchers to combine, analyze, and predict human behavior

¹College of Information Studies, University of Maryland, College Park, College Park, Maryland, United States

²CUNY The Graduate Center, New York, New York, United States

³Data & Society Research Institute, New York, New York, United States

⁴University of California, Irvine, California, United States

⁵University of Colorado at Boulder, Boulder, Colorado, United States

⁶Marquette University, Milwaukee, Wisconsin, United States

Corresponding author:

Katie Shilton, College of Information Studies, University of Maryland, College Park, College Park, Maryland, United States.
Email: kshilton@umd.edu



using multiple, partial, and disconnected datasets. Pervasive data research commonly collects such data through indirect means via partnerships with platforms, or through purchasing or scraping digital data. Even when researchers use direct means of data collection, they may have opportunities to repurpose or decontextualize consented data in ways the subject is not able to predict. The availability of pervasive digital data has destabilized the ethical relationship between researchers' methods for data collection and research subjects' autonomy to control their participation in research. Though many pervasive data studies have been uncontroversial, too frequently, participants react with alarm (Hallinan et al., 2020; Zimmer, 2018).

Existing institutional backstops designed to support public trust in human subject research have not met the challenge of establishing trustworthy pervasive data research practices. A recent column in *Science Magazine* pointed out the dearth of ethics guidance for big data research, noting: "the field has failed to fully articulate clear principles and mechanisms for collecting and analyzing digital data about people while minimizing the potential for harm" (Lazer et al., 2020: 1061). Although academic researchers have access to institutional guidance mechanisms in the form of ethics review boards such as Institutional Review Boards (IRBs) and Research Ethics Boards (REBs), and industry researchers are increasingly also setting up or accessing similar structures (Bowser and Tsai, 2015; Jackman and Kanerva, 2016) these boards frequently interpret data gathered through scraping or purchase as exempt from informed consent requirements (Nebeker et al., 2017a; Vitak et al., 2017). Their judgment rests on the assumption that such data are public and that no new risks to participants can come from the analysis of public data (Metcalf and Crawford, 2016)—an assumption proven untrue in the era of data science (Keegan and Matias, 2015; Rubinstein and Hartzog, 2016).

Pervasive data research is not the first field confronted with the shortcomings of traditional institutional ethics regulation. Ethnographic researchers have grappled with research ethics both within and beyond the framework of institutional review (Davies, 2012). Questions of whether and how to make the ethnographer's presence known to research subjects have been long debated in the literature (Bernard, 2006). Pervasive data research has more in common with ethnography than is immediately obvious. The instruments are different—human senses instead of digital sensors, individual sensemaking instead of algorithmic pattern-matching—but both forms of research rely on integration and interpretation of multiple data streams, and both require judgment about what features of a context are relevant for making meaning. The ethical challenges of research with pervasive data—the richly personal nature, the emphasis on observation, integration of multiple data types, and the drawing of inferences and conclusions based on patterns—are the same challenges that can be found in

the world of ethnography and participant observation. We are not the first to make this comparison; for example, Muller et al. (2016) discuss epistemic and ontological parallels between big data-based machine learning and grounded theory. However, here we explicitly use ethnographic research ethics as a guide for responding to challenges of awareness, power, and mistrust in pervasive data research. Ethnographers have deep experience in building trust with research subjects, as well as trust in the appropriateness and acceptability of their research—sometimes outside of or in conflict with the primary institutions and expectations of research ethics. Data scientists can use this experience, and the practices of ethnographic intervention, to help define trustworthy practice for pervasive data research.

We argue that pervasive data researchers must, like ethnographers, grapple with challenges of research subject awareness and acceptance as well as the appropriateness of their research, especially given data science's relationship to corporations, governments, and other sites of institutional domination. However, pervasive data research faces a challenge unaddressed by ethnographic tools. Ethnographers must explicitly negotiate awareness and power with their participants because of the physical embodiment their research requires, building trust along the way. But the ease of disembodied digital data gathering flattens the institutional structures traditionally relied on for building trust with research subjects. The challenge for pervasive data research is not only to center discussions of awareness and power in its research practices, but also to dig out from this disembodiment: to find ways to excavate and retexture modes of trust building.

This paper proceeds as follows. First, we detail evidence that many data subjects are largely unaware of the research uses of their digital communications and actions, and when they become aware, too frequently express unhappiness and alarm. We then introduce ways that ethnographic and participant-observation research have dealt with ethical challenges of research awareness as well as representational justice issues that stem from the power dynamics between researchers and research participants. Finally, we adapt those lessons for pervasive data and outline a foundation for trustworthy pervasive data research by engaging researchers in 1) rebuilding participant awareness and 2) excavating explicit considerations of power beyond traditional research ethics concerns.

2. The trustworthiness challenge

Trust and trustworthiness are complex constructs in the sociological, anthropological, and ethical literature, with debates over both the function and mechanics of trust and trustworthiness. In sociology, trust is often thought of as a construct necessary to deal with complexity and complex decision-making (Luhmann, 2017). Trust allows people to cooperate toward common goals, to pursue disparate

goals through partnerships, and to collectively manage uncertainty. Adapted for social sciences research, then, trust enables necessary forms of participation between researchers, individuals, and groups to foster the production of knowledge while managing the possible risks of entering into such partnerships.

Trust is a central problem in pervasive data research because of the methodology's reliance on datafication. Van Dijck (2014) argues that researchers should be wary of datafication as a core method for studying human behavior because the very possibility of datafication relies on fraught and brittle institutional trust. Data subjects may be willing to trade their behavioral data to corporate digital platforms in exchange for services, but that does not give researchers a just claim to that data or a good reason to expect the data to be representative of the underlying phenomena they seek to study. And because data science frequently contributes to automated decision-making, researchers have obligations to consider not only the potential impacts of new knowledge, but whether systems built with that knowledge would be trustworthy.

How might pervasive data researchers act in trustworthy ways in such a fraught environment? Trustworthiness is a form of right action (ethics) that emphasizes our duties and promises to other people (Tullberg, 2008). Our commitments to others can be made more credible (providing a reliable basis for the trust of others) through assurance mechanisms ranging from interpersonal dynamics to institutional constraints such as social norms and law (Hardin, 1996).

For researchers, commitments to research subjects have been defined for decades by principles put forth by the 1979 Belmont Report: respect for persons, beneficence, and justice. Belmont shaped assurance through policy constraints placed upon researchers in the U.S., Canada, Australia, and Europe. For example, in the U.S., trustworthy research practice is codified in the U.S. Common Rule, which interprets respect for persons as meaning "that subjects, to the degree that they are capable, be given the opportunity to choose what shall or shall not happen to them" (Office of the Secretary of The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). While necessary for trustworthiness, the principles of the Belmont Report and the Common Rule are not exhaustive of the scope of commitments and considerations that researchers ought to consider with respect to human research subjects, particularly for social scientists. As discussed below, the Belmont principles were designed to balance conflicting duties of care for physician-researchers, but social science researchers have oriented their approaches to trustworthiness around considerations of awareness and power. Therefore, social science researchers have had to come to terms with their own disciplines' histories of exploitation, extraction, and co-optation, as well as

their positionality as they foster relationships with research subjects (Sultana, 2007).

Trustworthy practice for pervasive data research—ensuring that researchers meet commitments like respect for persons, beneficence, and justice—is problematized by the ecosystem where digital research takes place. Research participants routinely deny knowledge of widespread research conducted with digital data and express that, while they might be willing to participate in digital data research, they expect to be asked for consent (Fiesler and Proferes, 2018; Gilbert et al., 2021; Hudson and Bruckman, 2004). However, a commitment to respect for persons, defined as self-determination for participants in the Common Rule, is difficult to ensure in pervasive data research. Informed consent is not always logistically or philosophically appropriate for research in the big data age. Logistically, there is now a large amount of data about people available online. Securing individual consent to use this data would be incredibly challenging—if not impossible—where individual identity was knowable, and arguably unethical where doing so would require collecting even more personal data (Ioannidis, 2013).

Philosophically, informed consent for pervasive data research suffers from a number of problems. Metcalf and Crawford (2016) point out that codes of informed consent were established specifically to govern physician-researchers, who balance the broad social interest in research results with their individual duty of care for a patient. The procedures and norms IRBs use to generate trust operate with an unstated assumption that these social conditions hold for all types of research. However, the trust relationships between computational social scientists, data scientists, and the public seldom conform to the social conditions that hold between physician-researchers and research subjects. The norms of pervasive data researchers (unlike, say, the norms of ethnographers) do not currently require preexisting, personal, or even explicitly declared relationships with the communities they study to collect data. The typical scale of data science manifests in numerous ethical and epistemic challenges for understanding the ethical interests of data subjects that extend beyond matters of logistics (Hanna and Park, 2020). Finally, Richards and Hartzog (2017) identify "pathologies" of consent resulting from overuse in the digital age. They argue that consent works best when it is given infrequently, when the harms are visceral and easily imagined, and when the stakes of a decision are significant. Pervasive data research meets some, but not all, of these standards. Explicit, informed consent to research participation happens infrequently for many participants. However, the harms of data research are rarely visceral or easily imagined. And it is unclear to what degree individuals consider the stakes of participating in research. Therefore, while it is unclear whether informed consent is philosophically the right mechanism to navigate the relationship between data scientists and data subjects, it is the case that many of the norms and mechanisms that

other forms of research use to achieve informed consent do not translate well to pervasive data research.

Challenges to defining trustworthy practices for pervasive data research extend beyond consent and echo larger social concerns with the power and social impacts of datafication. People are increasingly aware of—and alarmed by—the prevalence of datafication of their digital communications and activities (Auxier et al., 2019; Beninger, 2017; Dubois et al., 2020; Fiesler and Proferes, 2018; Golder et al., 2017; Gruzd and Mai, 2020; Hallinan et al., 2020; Hudson and Bruckman, 2004). For example, Hallinan et al. (2020) examined public reaction to Facebook’s emotional contagion study. One of their findings was that commenters objected to the idea of “living in a lab” or being studied without their awareness. Research also indicates that unwillingness to participate in pervasive data research is a larger concern among marginalized communities, where issues range from fear of surveillance and deportation (Nebeker et al., 2017b) to concerns that deployed technologies will fail to represent the needs and realities of user communities (Winchester, 2018) and to unwanted amplification of content or communities (Dym and Fiesler, 2020). Moreover, as Hoffman and Jonas (2017) point out, the costs of online participation are unequally borne by women and people of color, obligating researchers to consider the differential needs of vulnerable data subjects. Both distrust in platforms and concern for the uneven risks of surveillant research methods signal challenges of social power—who bears risk, and how partnerships with platforms shape that risk—that researchers must navigate.

Alarm over digital datafication was not (primarily) created by researchers. Influential works like Zuboff; (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* have introduced a broad public to the ways companies and governments use records of online activity to learn about, influence, and predict behavior. But though public alarm is largely a response to corporate data practices, data scientists are directly impacted. Hallinan et al. (2020) found that people’s angry reactions to the emotional contagion study were deeply bound up in opinions of the platform, Facebook, as much as or perhaps more so than the research itself. Public distrust in the platform bled into public distrust in the research.

Parallel to the growing public alarm about datafication has been a restriction of research access to some forms of pervasive data by social media platforms. As Tromble (2021) traces, platforms have moved away from open API access for researchers to narrower, more careful data access efforts enabled by tools like differential privacy (King and Persily, 2020). Tromble (2021) characterizes the “post-API” era as an opportunity for reflection by pervasive data researchers on the rigor and ethics of their data practices.

Data scientists are not alone in facing challenges of defining trustworthy research practice, coping with a

research ethics governance infrastructure misaligned with their methods and approaches, or responding to participant worries about invasive techniques and complex power dynamics. We believe there is useful instruction in the history of another research methodology that has struggled with trustworthiness: ethnography.

3. Trustworthy ethnographic research

In a classic ethnographic methods textbook, Spradley writes:

No matter how unobtrusive, ethnographic research always pries into the lives of informants. Participant-observation represents a powerful tool for invading other people’s way of life. It reveals information that can be used to *affirm* their rights, interests, and sensitivities or to *violate* them (1980: 22).

Substitute the words “pervasive data” and the concern is the same. Reflecting the first and foremost on whether data use will affirm the *rights, interests, and sensitivities* of the people it documents—or alternatively violate those interests—should be a first-order concern for trustworthy pervasive data research. Importantly, this ethical commitment leads with evaluative judgment about the purpose and effects of the research, not with a procedural determination of the status of the data.

This dedication to the rights and interests of the subjects of participant-observation comes from a long and painful history of the use of ethnographic methods. The history of ethnographic research is also one of colonialism. The earliest ethnographies were conducted by American and European academics through fieldwork among indigenous peoples, and this practice continued well into the 1980s (Thrift, 2003). The power-laden character of the colonial encounter in ethnographic fieldwork was recognized as a challenge from early on, as ethnographers were acutely aware of how they were seen by their informants as extensions of colonial government (Stocking, 1991). But as postcolonial movements took hold in academia, and particularly in anthropology (Asad, 1995), scholars began to grapple with, as Thrift describes, “whether it was possible to have encounters with others which were not inevitably, in some sense, colonial in form and content and had some genuine ethical weight” (2003: 107). A recentering of the rights of, and obligations to, research subjects was the result of this field-wide reckoning. Today, ethnographers are expected to reflect on their power as researchers, as well as on what they are taking from the communities they study. As Fine and Weis write, “Researchers can no longer afford to collect information on communities without that information benefiting those communities in their struggles for equity, participation, and representation” (1996: 293–294).

Ethical concerns about early ethnographies were not limited to colonialism but also extended to other historically disenfranchised groups. One of the most famous examples of controversy over ethical issues in ethnography surrounded Laud Humphreys' 1968 dissertation and later book titled *Tearoom Trade: Impersonal Sex in Public Places* (Humphreys, 1975). In this study, Humphreys conducted observations of male–male sexual encounters in public restrooms (“tearooms”). While the study had a positive impact on public perception of homosexuality, it also generated a significant amount of ethical controversy. In collecting his data, Humphreys hid the fact that he made research data from both sexual activity and participant identities. He disguised his identity and purpose as a researcher, did not obtain consent from his participants (although there were no standards for informed consent in research at the time), and used license plate numbers to discover the identities of the otherwise anonymous men from the restrooms. Some faculty in the Sociology Department at Washington University (where Humphreys received his PhD) were outraged and called for rescinding Humphreys' degree (Sieber, 1982: 3). Other scholars have defended the study, arguing that Humphreys “did not violate the deeper ethical and social concerns of sociology” (Lenza, 2004: 23). A widely-used sociology textbook claims that *Tearoom Trade* is “still debated, and it probably always will be, because it stirs emotions and involves ethical issues people disagree about” (Babbie, 2007: 74).

While the ethical merits of the study are debated, it is clear that *Tearoom Trade* negatively impacted *trust* in sociological research. In a *Washington Post* commentary about the study, journalist Nicholas von Hoffman stated:

We're so preoccupied with defending our privacy against insurance investigators, dope sleuths, counterespionage men, divorce detectives and credit checkers, that we overlook the social scientists behind the hunting blinds who're also peeping into what we thought were our most private and secret lives. But they are there, studying us, taking notes, getting to know us, as indifferent as everybody else to the feeling that to be a complete human involves having an aspect of ourselves that's unknown (1970: B1).

In this telling, the social scientist is not only invisible (behind hunting blinds), but also uncaring and fetishistic, part of a discipline that has a “peculiar taste for nosing around oddballs.” Von Hoffman concludes that even if Humphreys' motives were good, “no information is valuable enough to obtain it by nipping away at personal liberty” (1970: B1).

As research ethics legislation began to be codified internationally, researchers' growing engagement with the ethics of ethnographic fieldwork collided with university review. Despite ethnographers recognizing the need to build trust with both the communities they studied and

the institutions that employed them, many found the IRB process to be an obstacle to be overcome rather than a trust-building opportunity. The difficulties ethnographers faced in IRB review closely mirror the difficulties that pervasive data researchers now face with institutional review mechanisms. In Patricia A. Marshall's telling of this history, the primary challenges for ethnographers were “first, professional competency of IRBs to evaluate anthropological protocols; and second, applications of requirements for informed consent” (2003: 272). Without experience on ethnographic methods incorporated into IRBs, relatively low-risk research proposals were more likely to be elevated to the highest level of scrutiny, Marshall reports. And “the legalistic rendering of consent models used by most IRBs fail[ed] to recognize the social construction of informed consent as an act of communication” (2003: 274).

In response to the difficulties IRBs posed for ethnographers, Marshall offered recommendations that have since been adopted within fields such as anthropology and have eased (if not entirely resolved) the difficulties she identified. Marshall's recommendations included calls for representation of ethnographers on IRBs and outreach to IRBs from ethnographers to communicate ethnographic best practices. She also recommended robust documentation of research protocols for informed consent within the discipline, educating policymakers about the relevant challenges to informed consent for ethnographic fieldwork, and education within disciplines to include ethical guidelines in methodological training. Lastly, she called for research into how university review boards evaluate research proposals, observing that “[i]nformation on the decision-making processes used by IRBs in approving or rejecting a research proposal would be useful” (Marshall, 2003: 280) for crafting more effective and ethical ethnographic fieldwork protocols.

Ethnographers have developed a suite of techniques to gain the trust of participants. First, ethnographers must gain *entrée*: the permission of participants to be in their space and lives. As part of gaining *entrée*, ethnographers typically communicate their research objectives to participants and engage their participants in ongoing conversations about the research in progress. Another important facet of gaining *entrée* or permission to conduct the research is ensuring that participants receive something meaningful for their participation. Spradley's textbook recommends that “every ethnographic research project should, to some extent, include a dialogue with informants to explore ways in which the study can be useful to informants” (1980: 22–23).

Entrée is only the beginning of awareness in ethnographic projects. Ethnographers frequently encourage research participants to read parts or all of their analysis. *Participant checking* is the practice of sharing everything from raw data (quotations and anecdotes) to analyses and conclusions with project participants. The ethnographer does this to check the validity of their observations and findings, and to spur

potentially interesting new conversations about the research itself. Participant checking acknowledges that research participants also have expertise and helps to increase their participation in new knowledge creation.

Some ethnographic traditions move beyond *entrée* and participant checking to *collaborative ethnography*, which departs from the traditional models of research that have a clear separation between the subject and the scholar to involving people about whom the research investigates from start to finish, including in the creation of the research goals, data collection, analysis, authorship, and dissemination (Fluehr-Lobban, 2008; Lassiter, 2005). In this way, collaborative ethnography is meant to provide a more ethical approach to research by conducting research *for* or *with* people rather than *on* them (Fluehr-Lobban, 2008). The motivations driving collaborative ethnography, as well as the solutions, are also found in other methodological paradigms, such as action research (Khanlou and Peter, 2005). Already used in computing fields such as Human–Computer Interaction, action research repositions researchers as facilitators whose role is to provide expertise while working with a community towards a social action (Hayes, 2014).

These methodological interventions accompanied a broader, deliberate, decades-long theoretical and political shift that decoupled ethnography from the levers of colonial, state, and corporate power (Alkhatib, 2019). This shift consisted of several key components. One was a series of field-defining self-critical investigations of how the discipline had functioned as an agent of these powers (Asad, 1973). Another was the gradual growth of broader demographic representation within the discipline beyond historically white male cohorts (Patterson, 2020). Yet another component was a shift in topics of study, answering the call for *studying up* (Nader, 1969): performing ethnographic research on groups more (or at least equally), rather than less, socially powerful than the researcher. Studying politicians, judges, scientists, corporate executives and managers, and white-collar criminals can mitigate some of the concerns about whether data collection serves the needs of the community first, because society values *accountability* of its powerful people and institutions. More recently, professional anthropological associations have taken steps to declare that ethnography undertaken in the service of state security and military apparatuses, particularly those like the Human Terrain System (Price, 2011), explicitly violate the associations' Code of Ethics (AAA, 2007, 2012).

4. Excavating awareness and power in pervasive data research

Data collection, once the province of researchers, is now dominated by companies and governments. Growing distrust in big

data research hinges on the fact that people increasingly realize how vulnerable the datafication of their lives makes them both to commercial platforms and governments, which use pervasive data to sell and surveil, categorize and control, as well as to discipline and discriminate. By employing pervasive data as a tool for research, data scientists participate in this legacy, and—much like ethnographers grappling with the extractive, colonial legacy of their methodology—must take specific action to address their place in the entangled social problems of digital data analysis. This means data scientists should be thinking about all of the powerful corporate, state, and societal forces entangled in big data. However, data scientists must do so in an environment in which traditional structures that support trustworthiness—clear practice norms or guidelines, direct interaction with research subjects, approval by ethics review boards, and distinctions between academic and commercial benefits—are absent or much less visible. How can we excavate structures to support trustworthiness that have been flattened in pervasive data research?

To retexturize this flattened landscape, we recommend that data scientists probe appropriateness and complex potential harms using two lenses directly inspired by ethnography. First, data scientists can learn from ethnographers' contextual sensitivity and experience helping research subjects interpret research participation by reflecting on *awareness*. How can data scientists resist the unawareness so endemic to big data collection and use, and make their data use practices known to, and acceptable to, the populations that they are studying? Second, we argue that data scientists must reflect upon *power* by considering the potential representational harms of their research practices. This requires both reflecting on their own subjectivity as researchers and centering the rights, interests, and sensitivities of the people being studied (Abu-Lughod, 1996; Clifford and Marcus, 2010). We provide approaches to support such reflection below.

Awareness

Supporting awareness of datafication begins with reflection on the nature of the communications or traces being used for research. We map awareness of pervasive data to two spectra based on how digital traces are created: traces created in private to public settings, and traces created by intentional to automatic means. Private, intentional data trails are “secrets”: for example, texts to a spouse or family photos. People are aware they are creating communications or documentation, and also make conscious efforts not to share them widely. Public, intentional data trails are “broadcasts”: for example, tweets. People are aware they are creating communication or documentation and purposefully share them widely, even if they may not understand the extent of their reach (Proferes, 2017). Private, automatic data trails are “espionage”: communications or documentation created without human intervention or awareness and not widely shared. Examples include the

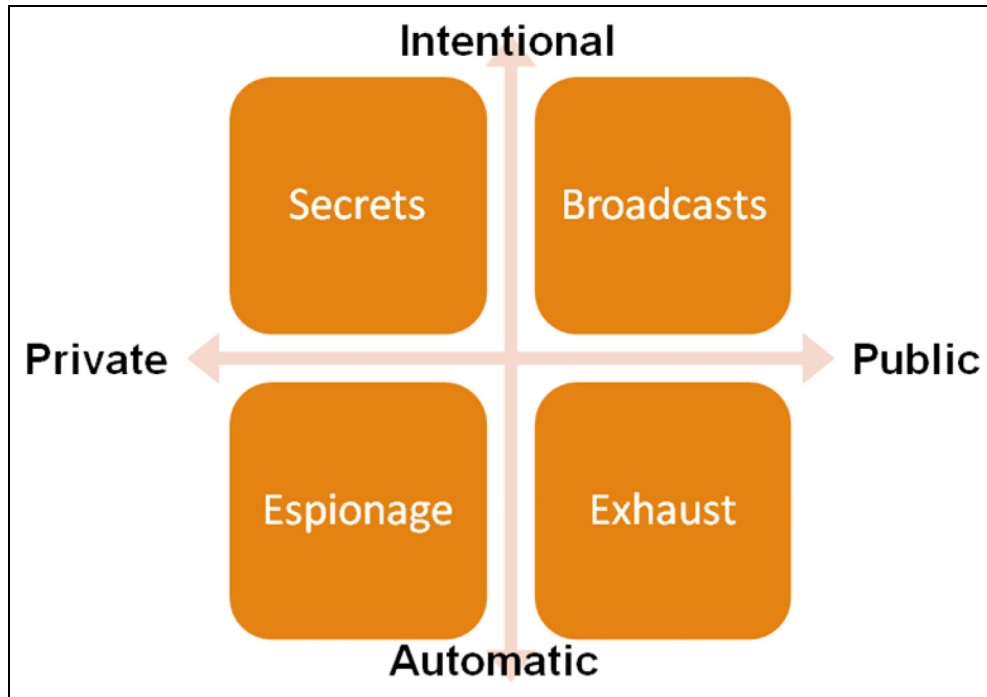


Figure 1. Spectra of data awareness.

data collected by a DVR, smart fridge, or thermostat. “Espionage” data can also include geolocation or telemetry data collected as part of the normal functioning of devices like smartphones. Though users may be aware of functions that require such documentation, they may not know the extent to which it is being collected (Hannay and Baatard, 2011). Finally, public automatic data can be thought of as “exhaust”: documentation captured in public that individuals are not aware they are putting out into the world. Examples are CCTV camera recordings or satellite images of a home.

We recommend that researchers reflect on where their data-gathering methods fall on the private/public and automatic/intentional spectra (Figure 1) and use this reflection as a guide for considering both awareness and power implications of their research. Researchers using “broadcasts” should consider that while participants were likely aware of creating a communication, they may not be aware of its potential for use in research. Crucially, using “public” data does not automatically relieve the researcher from considerations of participant awareness, because awareness of creation is not necessarily awareness of research use. Data scientists using “espionage” have an even higher bar to clear, as they need to make participants aware of both the existence of a data trace as well as its use for research.

Meaningful informed consent is one standard for raising data subjects’ awareness of research. But for pervasive data researchers who can’t secure meaningful consent because of scale, pervasiveness, or other issues, adaptations of entrée and participant checking should evolve with the field. As

a starting point, they might include website pop-ups, such as those used as part of GDPR notification requirements, that explain the research, risk to participants, and allow potential subjects to opt out. Another way to build awareness is to increase research subjects’ participation in the research. Increasing participation will require researchers to think about what participation looks like for the community or population they study (Sloane et al., 2020). Defining a community or population encourages researchers to resist universalizing data science findings, instead paying particular attention to who is and who is not included in the data under analysis (Costanza-Chock, 2018; Hargittai, 2015). Participatory action research (Khanlou and Peter, 2005) has grappled extensively with questions of participation, motivation, and accessibility, and can guide data scientists on challenging questions such as how to define a community, how to structure participation, and how to ensure representation of stakeholders across a community.

Power

Pervasive data researchers and the institutions that support them should also explicitly consider power relations and representational justice: they must consider the appropriateness of converting digital traces into research data, much as ethnographers have learned to consider when and if they should exert their power to transform particular groups into research subjects. However, just as with awareness, pervasive data researchers face an additional challenge. Ethnographers often experience both the power and

vulnerability of their participants because of the embodied, affective experiences of being in the field (Ceglowski, 2002). Researchers who have collected or acquired datasets without direct contact with participants do not have this embodied experience as a reminder or cue.

Therefore, pervasive data researchers must remind themselves to be reflexive about the power they hold and they should try to reflect on how their research question development, data creation or data gathering techniques, data analysis, and writing benefit from their power over data subjects, as well as the context of their research and their own subjectivity (Hegelund, 2005). They should consider whether it is appropriate to make a given community, stakeholder group, or population more vulnerable either by creating new forms of data (which may be used by other parties to increase their vulnerability) or through secondary uses of data: by making research data out of traces or communications created for other purposes. This consideration might involve spending time (virtually or physically) in a community to understand their norms, collaborating with a community to serve their needs, or speaking to community gatekeepers to understand specific harms—for example, amplifying content beyond its intended audience (Dym and Fiesler, 2020).

Using pervasive data for research increases the vulnerability of the people included (and potentially other people like those included), whether by amplifying their behaviors and beliefs, showing new connections or inferences between their activities and habits, or applying categories or labels to their actions. Scholars are increasingly developing approaches to help researchers think through implications and harms of datafication, such as the Omidyar Network's *Ethical Explorer Pack* (Artefact Group, n.d.) and Wong et al.'s (2017) privacy by design workbooks. And statements of potential vulnerabilities, harms, and biases are increasingly required in technical research communities (Gibney, 2020). It is important that scientists continue to use pervasive data to study those who need increased representation in knowledge (e.g. people with rare diseases, groups otherwise marginalized in research). Even so, pervasive data research should emphasize the standard drawn from political representation movements and disability activism: nothing about us without us (Charlton, 2004). Writing, "Anytime a research team is studying or working with a marginalized group, they must acknowledge an additional burden of representation," Ogbonnaya-Ogburu et al. (2020: 9-10) point out both epistemological and ethical problems with research on marginalized groups when those groups are not represented among the research team. Too often, as Patricia Hill Collins writes: "Black women and similarly marginalized groups were the 'facts' or data for manipulation within other peoples' preconceived frameworks about social problems" (2015: 2349). Using pervasive data to study populations without significant representation from those group

risks erasing the standpoints and lived experiences of the people behind that data.

To excavate those viewpoints, we recommend that researchers reflect on their power relative to the people they are studying. A clear theory of power is necessary for pervasive data research because there is a strong ethical argument for studying very *powerful* groups with less burden for awareness or participation. Such groups might include people with significant social or political power (legislators, technology companies, and architects of the financial system) for whom public accountability is important. Similarly, there is a strong justice argument to be made for using pervasive data to do research on those perpetrating violence, hatred, and bigotry. As Massanari (2018) argues, extremist groups complicate and shift the power dynamic between researchers and research subjects by using surveillance, threats, and harassment to intimidate researchers. To safely study groups exerting power through violence or the threat of violence (and therefore hold such groups publicly accountable), researchers may need to use digital traces without awareness.

As part of systematically examining power in pervasive data research, researchers should reflect on their relationships with platforms that create, collect, mediate, and disseminate pervasive data. The relationship between researchers and platforms is fraught with both trust issues and institutional privilege (Cooky et al., 2018). But cultivating trustworthy pervasive data research must include engaging with the platforms towards ethical new knowledge creation: pervasive data research cannot be confined to the secrecy (and benefit) of industry. Platforms, like universities, can contribute to institutional trust in pervasive data research and researchers, but only if their relationship to researchers is open and transparent. Pervasive data researchers partnering with corporations for data access should reflect on how and whether they can provide sufficient transparency into the researcher/platform relationship. They should also be aware of, and transparent about, the limitations that platform dynamics put on the representativeness and meaning-making potential of their data.

Researchers should also reflect on their data's relationships to, and potential interest to, governments. State power is increasingly exercised through purchase or subpoena of data created by platforms. As researchers contribute to what can be known about data subjects, they ought to consider how that knowledge may serve state power. This is particularly applicable to domains like criminal justice and immigration enforcement, but research ought to consider how all data science work might bolster the already powerful—a process that anthropology similarly had to undertake (Alkhatib, 2019).

Of course, determining one's power relative to research subjects, platforms, and state actors is a complex process, and we realize that such reflection is a difficult task for many without a background in theories of power.

However, there are numerous helpful frameworks for thinking through issues of power adapted specifically for digital data research, including anti-essentialism (Neyland, 2016), feminism (D'Ignazio and Klein, 2020; Hoffman and Jonas, 2017), anti-racism (Benjamin, 2019; Hanna et al., 2020), anti-colonialism (Dourish and Mainwaring, 2012), and queer theory (Brim and Ghaziani, 2016). These frameworks can provide concrete guidance to researchers considering how their protocols might unevenly subject participants to increased vulnerability. And many pervasive data researchers are already successfully grappling with such frameworks. For example, Blodgett et al. (2020) have shown how discussions of “bias” in natural language processing are largely discussions of power, particularly the relationship between language and existing social hierarchies. Barabas et al. (2020) have called for studying up as a technique to improve algorithmic fairness, and Mathur et al. (2020) have used pervasive data to analyze corporate and political deceptive practices. Mohamed et al. (2020) have grappled with how to embed decolonial theories in AI practice. Kaleidoscope (2019) is a “positionality aware” machine learning project that seeks to move discussions of researcher and data context to the center of data research. And the “Tech Won’t Build It” movement has moved discussions of power and refusal from academic research to corporate practice (Science for the People, 2018). We hope that more leaders in pervasive data research will follow these examples.

5. Conclusion

Reflections on awareness and power are not required of ethnographers—or data scientists—by institutional regulatory structures like IRBs and REBs. Within ethnographic research, adherence to these reflections has become bound up with definitions of ethical research, good methods, and good practice. The community of ethnographic researchers holds each other accountable for reflection on awareness and power, performed through techniques such as positionality statements and reflexivity in writing. Ethnographic researchers have also committed themselves (however imperfectly) to broadening the scope of participation in the discipline and have demonstrated a willingness to explicitly disavow work that conflicts with their professional organizations’ codes of ethics. We recognize that awareness and power are not the only ethical considerations in pervasive data research, but reflecting on these issues is a prerequisite to more trustworthy forms of research.

Both individual researchers and the myriad of professional organizations that support pervasive data research can contribute to a more trustworthy research community. We hope that research leaders will lead reflection on awareness and power with their students and collaborators, and that journal and conference reviewers will look for these reflections in the methods sections of papers. We also hope that professional organizations will establish

professional norms and codes for trustworthy pervasive data research, as anthropologists and sociologists have done. While some Internet- and AI-focused research organizations have been leaders in this area (Castelvecchi, 2021; Franzke et al., 2020), with the increased adoption of data science methods into the standard research repertoires of many fields, there is a need for traditional professional organizations to join codification efforts. While norms and codes are insufficient on their own, they are important to set standards of professional conduct, disseminate appropriate research practices, and resolve ambiguities and disputes over what constitutes appropriate research practice. Professional standards are also important mechanisms for learning from past experience and enshrining such lessons in public articulations of values, ethics, and norms. The history of research ethics is, in many ways and for many disciplines, comprised of the scar tissue that has grown over past controversies. The state of research ethics has gradually improved as researchers have examined their past and put in place interventions to ensure such controversies are not replicated in the future.

Digital device users are already quite vulnerable as their traces are turned into data by researchers, corporations, and governments. We hope that pervasive data research will embrace reflections on how to decrease this vulnerability, and will hold each other to higher ethical standards, particularly with respect to openness with research subjects and reflexivity on power and impact. As social computing, computational social science, health, natural language processing, and other pervasive data researchers consider their methods and sampling strategy, they should also ask themselves: are my research practices trustworthy? Critically, trustworthiness is a *means*—a process—not just an *end*. Whether pervasive data research ultimately can gain participants’ trust is not entirely up to researchers. What is up to us is trustworthy practice. The datafication of human activities for new knowledge creation has always demanded increased attention to questions awareness and justice. Pervasive data research is simply the latest research practice to need to redefine what, exactly, those principles look like in practice.

Acknowledgements

This article was a product of the PERVADE project, an NSF-funded collaboration focused on empirical questions in big data ethics. Many thanks to Emily Dacquist for logistical support on this article and throughout the project. Thanks also to the anonymous reviewers who substantially strengthened our arguments and presentation.


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation's Division of Information and Intelligent Systems awards 1704369, 1704425, 1947754, 1704598, and 1704303.

ORCID iDs

Katie Shilton  <https://orcid.org/0000-0003-1816-6140>

Matthew J. Bietz  <https://orcid.org/0000-0002-7782-3887>

Jessica Vitak  <https://orcid.org/0000-0001-9362-9032>

Michael Zimmer  <https://orcid.org/0000-0003-4229-4847>

References

- AAA (2007) *American Anthropological Association's Executive Board Statement on the Human Terrain System Project*. Arlington, VA: American Anthropological Association.
- AAA. (2012) *Full Text of the 2012 Ethics Statement*. American Anthropological Association. Available at: <http://ethics.americananthro.org/category/statement/> (accessed 14 January 2021).
- Abu-Lughod L (1996) Writing against culture. In: Fox RG (ed) *Recapturing Anthropology: Working in the Present*. School of American Research, 14: 137–162.
- Alkhatib A (2019) “Anthropological/Artificial Intelligence & the HAI.” Ali Alkhatib Blog. March 26, 2019. Accessed July 28, 2021. <https://ali-alkhatib.com/blog/anthropological-intelligence>.
- Artefact Group (n.d.) The ethical explorer pack. Available at: <https://www.artefactgroup.com/case-studies/ethical-explorer-pack/> (accessed 8 June 2021).
- Asad T (1995) *Anthropology & the Colonial Encounter*. London, UK: Humanities Press.
- Auxier B, Rainie L, Anderson M, et al. (2019) Americans And privacy: Concerned, confused and feeling lack of control over their personal information. *Pew Research Center* 15, November: 1–62.
- Babbie E (2007) *The Practice of Social Research (Vol. 11th)*. Belmont, CA: Wadsworth.
- Barabas C, Doyle C, Rubinovitz J, et al. (2020) Studying Up: Reorienting the study of algorithmic fairness around issues of power. In: *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, 27–30 January 2020, pp.167–176. New York, NY: Association for Computing Machinery.
- Beninger K (2017) Social media users' views on the ethics of social media research. In: Sloan L and Quan-Haase A (eds) *The SAGE Handbook of Social Media Research Methods*. Los Angeles, CA: Sage, 57–73.
- Benjamin R (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*. Hoboken, NJ: John Wiley & Sons.
- Bernard HR (2006) *Research Methods in Anthropology: Qualitative And Quantitative Approaches*. Lanham, MD: Rowman Altamira.
- Blodgett SL, Barocas S, Daumé HIII, et al. (2020) Language (technology) is power: A critical survey of “bias” in NLP. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, 5–10 July 2020, pp. 5454–5476. Stroudsburg, PA: Association for Computational Linguistics. 10.18653/v1/2020.acl-main.485
- Bowser A and Tsai JY (2015) Supporting ethical web research: A new research ethics review. In: *Proceedings of the 24th International Conference on World Wide Web*, Florence, Italy, 18–22 May 2015, pp. 151–161. New York: Association for Computing Machinery. doi:10.1145/2736277.2741654
- Brim M and Ghaziani A (2016) Introduction: Queer methods. *Women's Studies Quarterly* 44(3/4): 14–27.
- Castelvecchi D (2021) Prestigious AI meeting takes steps to improve ethics of research. *Nature* 589: 12–13.
- Ceglowski D (2002) Research as relationship. In: Denzin NK and Lincoln YS (eds) *The Qualitative Inquiry Reader*. Thousand Oaks, CA: Sage Publications, 5–24.
- Charlton JI (2004) *Nothing About Us Without Us: Disability, Oppression, and Empowerment*. Berkeley, CA: University of California Press.
- Clifford J and Marcus GE (2010) *Writing Culture: The Poetics and Politics of Ethnography* (Second Edition, 25th Anniversary). Berkeley, CA: University of California Press.
- Collins PH (2015) Intersectionality's definitional dilemmas. *Annual Review of Sociology* 41: 1–20.
- Cook C, Linabary JR and Corple DJ (2018) Navigating big data dilemmas: Feminist holistic reflexivity in social media research. *Big Data & Society* 5(2): 1–20.
- Costanza-Chock S (2018) Design justice, A.I., and escape from the matrix of domination. *Journal of Design and Science*. doi:10.21428/96c8d426.
- Cukier K and Mayer-Schoenberger V (2013) The rise of big data: How it's Changing the way we think about the world. *Foreign Affairs* 92(3): 28–40.
- Davies CA (2012) *Reflexive Ethnography: A Guide to Researching Selves and Others*, 2nd edition Abingdon, UK: Routledge.
- D'Ignazio C and Klein LF (2020) *Data Feminism*. Cambridge, MA: The MIT Press.
- Dourish P and Mainwaring SD (2012) Ubicomp's Colonial Impulse. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, Pittsburgh, Pennsylvania, 5–8 September 2012, pp. 133–142. New York, NY: Association for Computing Machinery. doi:10.1145/2370216.2370238
- Dubois E, Gruzd A and Jacobson J (2020) Journalists' use of social media to infer public opinion: The citizens' perspective. *Social Science Computer Review* 38(1): 57–74.
- Dym B and Fiesler C (2020) Ethical and privacy considerations for research using online fandom data. *Transformative Works and Cultures* 33.
- Fiesler C and Proferes N (2018) “Participant” perceptions of twitter research ethics. *Social Media+Society* 4(1): 1–14.
- Fine M and Weis L (1996) Writing the “wrongs” of fieldwork: Confronting our own research/writing dilemmas in urban ethnographies. *Qualitative Inquiry* 2(3): 251–274.
- Fluehr-Lobban C (2008) Collaborative anthropology as twenty-first-century ethical anthropology. *Collaborative Anthropologies* 1(1): 175–182.
- franzke as, Bechmann A, Zimmer M, et al. and the Association of Internet Researchers Ethics Working Group (2019) Internet research: Ethical guidelines 3.0. Report, Association of Internet Researchers, October. <https://aoir.org/reports/ethics3.pdf>
- Gibney E (2020) The battle for ethical AI at the world's Biggest machine-learning conference. *Nature* 577: 09.

- Gilbert SA, Vitak J and Shilton K (2021) Measuring Americans' comfort with research uses of their social media data. *Social Media+Society* 7(3): 1–13.
- Golder S, Ahmed S, Norman G, et al. (2017) Attitudes toward the ethics of research using social media: A systematic review. *Journal of Medical Internet Research* 19(6): e195.
- Gruzd A and Mai P (2020) Social media privacy in Canada 2020. Report, Scholars Portal Dataverse, November. doi:10.5683/SP2/PUWDYC.
- Hallinan B, Brubaker J and Fiesler C (2020) Unexpected expectations: Public reaction to the facebook emotional contagion study. *New Media & Society* 22(6): 1076–1094.
- Hanna A, Denton E, Smart A, et al. (2020) Towards a critical race methodology in algorithmic fairness. In: *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, 27–30 January 2020, pp.501–512. New York, NY: Association for Computing Machinery. doi:10.1145/3351095.3372826.
- Hanna A and Park TM (2020) Against scale: Provocations and resistances to scale thinking. *Computers and Society*. November: 2. <https://arxiv.org/abs/2010.08850>.
- Hannay P and Baatard G (2011) Geointelligence: Data mining locational social media content for profiling and information gathering. In: *2nd International Cyber Resilience Conference*, Perth, Western Australia, 1–2 August 2011, pp. 29–37. Perth, Western Australia: Edith Cowan University. <https://ro.ecu.edu.au/icr/20>
- Hardin R (1996) Trustworthiness. *Ethics* 107(1): 26–42.
- Hargittai E (2015) Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science* 659(1): 63–76.
- Hayes GR (2014) Knowing by doing: Action research as an approach to HCI. In: Olson JS and Kellogg WA (eds) *Ways of Knowing in HCI*. Berlin, Germany: Springer, 49–68.
- Hegelund A (2005) Objectivity and subjectivity in the ethnographic method. *Qualitative Health Research* 15(5): 647–668.
- Hoffman AL and Jonas A (2017) Recasting justice for internet and online industry research ethics. In: Zimmer M and Kinder-Kurlanda K (eds) *Internet Research Ethics for the Social age*. Bern, Switzerland: Peter Lang, 3–18.
- Hudson JM and Bruckman A (2004) “Go away”: Participant objections to being studied and the ethics of chatroom research. *The Information Society* 20(2): 127–139.
- Humphreys L (1975) *Tearoom Trade: Impersonal Sex in Public Places*, 2nd edition Abingdon, UK: Routledge.
- Ioannidis JPA (2013) Informed consent, big data, and the oxymoron of research that is not neresearch. *The American Journal of Bioethics* 13(4): 40–42.
- Jackman M and Kanerva L (2016) Evolving the IRB: Building robust review for industry research. *Washington and Lee Law Review Online* 72(3): 442–457.
- Kaleidoscope (2019) Available at: <https://kaleidoscope.media.mit.edu/> (accessed 21 January 2021).
- Keegan BC and Matias JN (2015) Actually, it's About ethics in computational social science: A multi-party risk-benefit framework for online community research. *ArXiv:1511.06578 [Cs]*. <http://arxiv.org/abs/1511.06578>.
- Khanlou N and Peter E (2005) Participatory action research: Considerations for ethical review. *Social Science & Medicine* 60: 2333–2340.
- King G and Persily N (2020) Unprecedented Facebook URLs dataset now available for academic research through social science One. In: *Social Science One*. Harvard's Institute for Quantitative Social Science. Available at: <https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one> (accessed 7 June 2021).
- Lassiter LE (2005) *The Chicago Guide to Collaborative Ethnography*. Chicago, IL: The University of Chicago Press.
- Lazer DMJ, Pentland A, Watts DJ, et al. (2020) Computational social science: Obstacles and opportunities. *Science (New York, N Y)* 369(6507): 1060–1062.
- Lenza M (2004) Controversies surrounding Laud Humphreys' tearoom trades? An unsettling example of politics and power in methodological critiques. *International Journal of Sociology and Social Policy* 24: 20–31.
- Luhmann N (2017) *Trust and Power*. Cambridge, UK: Polity.
- Marshall PA (2003) Human subjects protections, institutional review boards, and cultural anthropological research. *Anthropological Quarterly* 76(2): 269–285.
- Massanari AL (2018) Rethinking research ethics, power, and the risk of visibility in the era of the “alt-right” gaze. *Social Media+Society* 4(2): –9.
- Mathur A, Wang A, Schwemmer C, et al. (2020) Manipulative tactics are the norm in political emails: Evidence from 100K emails from the 2020 U.S. election cycle. Epub ahead of print 5 October 2020. Available at: <http://64.225.26.49/assets/manipulative-political-emails-working-paper.pdf>.
- Metcalf J and Crawford K (2016) Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society* 3(1): 1–14.
- Mohamed S, Png MT and Isaac W (2020) Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33(4): 659–684.
- Muller M, Guha S, Baumer EPS, et al. (2016) Machine learning and grounded theory method: Convergence, divergence, and combination. In: *Proceedings of the 19th International Conference on Supporting Group Work*, Sanibel Island, Florida, 13–16 November 2016, pp.3–8. New York, NY: Association for Computing Machinery. doi:10.1145/2957276.2957280.
- Nader L (1969) Up the anthropologist: Perspectives gained from studying up. In: Hymes D (ed) *Reinventing Anthropology*. New York, NY: Random House, 1–28. <https://eric.ed.gov/?id=ED065375>.
- Nebeker C, Harlow J, Giacinto-Espinoza R, et al. (2017a) Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. *AJOB Empirical Bioethics* 8(4): 266–276.
- Nebeker C, Murray K, Holub C, et al. (2017b) Acceptance of mobile health in communities underrepresented in biomedical research: Barriers and ethical considerations for scientists. *JMIR MHealth and UHealth* 5(6): e87.
- Neyland D (2016) Bearing account-able witness to the ethical algorithmic system. *Science, Technology, & Human Values* 41(1): 50–76.
- Office of the Secretary of The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979). The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Department of Health, Education, and Welfare.

- Ogbonnaya-Ogburu IF, Smith ADR, To A, et al. (2020) Critical Race Theory for HCI. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, Hawaii, 25-30 April 2020, pp.1–16. New York, NY: Association for Computing Machinery. doi:10.1145/3313831.3376392.
- Patterson TC (2020) *A Social History of Anthropology in the United States*. Oxfordshire, UK: Taylor & Francis.
- Price DH (2011) *Weaponizing Anthropology: Social Science in Service of the Militarized State*. Stirling, UK: AK Press.
- Proferes N (2017) Information flow solipsism in an exploratory study of beliefs about twitter. *Social Media & Society* 2017: 1–17.
- Richards N and Hartzog W (2017) Trusting big data research. *DePaul Law Review* 66(2): 578–590.
- Rubinstein I and Hartzog W (2016) Anonymization and risk. *Washington Law Review* 91(2): 703–760. <http://papers.ssrn.com/abstract=2646185>.
- Science for the People (2018) Solidarity letter: Tech won't build it! Available at: <https://scienceforthepeople.org/2018/09/25/solidarity-letter-tech-wont-build-it/> (accessed 21 January 2021).
- Siebert JE (1982) *Ethical Dilemmas in Social Research. In The Ethics of Social Research: Surveys and Experiments*. Berlin, Germany: Springer-Verlag.
- Sloane M, Moss E, Awomolo O, et al. (2020) Participation is not a design fix for machine learning. In: *Proceedings of the 37th International Conference on Machine Learning*, Online, 13-18 July 2020, pp. 7. <http://arxiv.org/abs/2007.02423>
- Spradley JP (1980) *Participant Observation*. New York, NY: Holt, Rinehart and Winston.
- Stocking GW (1991) *Colonial Situations: Essays on the Contextualization of Ethnographic Knowledge* (Vol. 7). Madison, WI: The University of Wisconsin Press. <https://uwpress.wisc.edu/books/0339.htm>.
- Sultana F (2007) Reflexivity, positionality and participatory ethics: Negotiating fieldwork dilemmas in international research. *ACME: An International Journal for Critical Geographies* 6(3): 374–385.
- Thrift N (2003) Practising ethics. In: *Using Social Theory: Thinking Through Research*. Thousand Oaks, CA: Sage, 105–121.
- Tromble R (2021) Where have all the data gone? A critical reflection on academic digital research in the post-API age. *Social Media+Society* 7(1): –8.
- Tullberg J (2008) Trust—The importance of trustfulness versus trustworthiness. *The Journal of Socio-Economics* 37(5): 2059–2071.
- Van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- Vitak J, Proferes N, Shilton K, et al. (2017) Ethics regulation in social computing research: Examining the role of institutional review boards. *Journal of Empirical Research on Human Research Ethics* 12(5): 372–382.
- von Hoffman N (1970) Sociological snoopers. *The Washington Post*, B1.
- Winchester WWIII (2018) Afrofuturism, inclusion, and the design imagination. *Interactions* 25(2): 41–45.
- Wong RY, Mulligan DK, Van Wyk E, et al. (2017) Eliciting values reflections by engaging privacy futures using design workbooks. *Proceedings of the ACM on Human-Computer Interaction* 1: 1–26.
- Zimmer M (2018) Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media +Society* 4(2): 1–11.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for A Human Future at the New Frontier of Power*, 1st edition New York, NY: PublicAffairs.