

Ethics as Methods: Doing Ethics in the Era of Big Data Research—Introduction

Annette N Markham¹ , Katrin Tiidenberg^{1,2},
and Andrew Herman³

Social Media + Society
July–September 2018: 1–9
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2056305118784502
journals.sagepub.com/home/sms

Abstract

This is an introduction to the special issue of “Ethics as Methods: Doing Ethics in the Era of Big Data Research.” Building on a variety of theoretical paradigms (i.e., critical theory, [new] materialism, feminist ethics, theory of cultural techniques) and frameworks (i.e., contextual integrity, deflationary perspective, ethics of care), the Special Issue contributes specific cases and fine-grained conceptual distinctions to ongoing discussions about the ethics in data-driven research. In the second decade of the 21st century, a grand narrative is emerging that posits knowledge derived from data analytics as true, because of the objective qualities of data, their means of collection and analysis, and the sheer size of the data set. The by-product of this grand narrative is that the qualitative aspects of behavior and experience that form the data are diminished, and the human is removed from the process of analysis. This situates data science as a process of analysis performed by the tool, which obscures human decisions in the process. The scholars involved in this Special Issue problematize the assumptions and trends in big data research and point out the crisis in accountability that emerges from using such data to make societal interventions. Our collaborators offer a range of answers to the question of how to configure ethics through a methodological framework in the context of the prevalence of big data, neural networks, and automated, algorithmic governance of much of human sociality.

Keywords

ethics, method, social media

In recent years, we have witnessed several situations that raise new and old questions about ethics. Just as we wrap up this Special Issue, in fact, our news feeds unveil details of the latest case, where Cambridge Analytica, a large data mining and analysis firm, was able to access personal details of 50 million Facebook users without their direct permission or knowledge. This case is just the latest in a long list that illuminate the complications of ethics in social media practices (i.e., wide scale public shaming), data management (i.e., deliberate or accidental releases of private information), data slippage (i.e., moving from one context to another), technology design (i.e., search engine bias), and multiple other areas relevant for (social) life today.

The locus of responsibility and accountability for ethical design, behavior, and outcomes is difficult to ascertain. Every social media situation involves multiple moments, decisions, actions, and operations that can result in outcomes that have potential harm for people. A complex ecology of simultaneously functioning systems and entities with a variety of interests, power, and self-awareness operates underneath the

apparent seamlessness of our interfaces. The challenge of locating responsibility and accountability is exacerbated by the difficulty of determining with any clarity the relationship between action and consequence as well as between data and persons. This is not just a matter of finding the proverbial paper trail, as many of our ontological assumptions about the distinctions between data produced by people and people themselves are challenged. Indeed, one of the hallmarks of the era of “big data” is the abstraction and disarticulation of data about individuals whose activity in digital spaces is the source of the data. In other words, through the symbolic alchemy of algorithmic computation, there is a transmutation of the

¹Aarhus University, Denmark

²Tallinn University, Estonia

³Wilfrid Laurier University, Canada

Corresponding Author:

Katrin Tiidenberg, Baltic Film, Media, Arts and Communication School,
Tallinn University, Uus-Sadama 5, Tallinn, Harju 10120, Estonia.
Email: katrin.tiidenberg@gmail.com



person as a coherent representational entity into a constellation of data points abstracted from social context and lived experience. When our everyday actions, movements, and utterances are transformed into data points and traded on the stock market (Skeggs, 2017) it is clearly time to reframe what constitutes the boundaries of being.

At the epistemological level, we can ask, among other questions, what practices of knowing are we using to determine these boundaries? And what are the stakes behind our decisions? Katherine Hayles (2015), when talking about high-frequency trading, has pointed out that even if we do not all agree that non-human entities like algorithms or self-driving cars have the same level of agency as humans, we must address and evaluate them *as if* they do at the epistemological level, since the outcomes of their actions, functions, or calculations have serious impact.

The authors of this Special Issue have paid close attention to both the epistemological and ontological to address the axiological and methodological. They pose many responses to the question: How do we “do” ethics in this epoch, and what are its possible impacts? This specific stance about the relationship of ethics and methods is what is articulated in the title of this Special Issue. We position ethics as methods and, vice versa, methods as ethics. This deliberate conflation builds on Annette Markham’s (2003, 2006) idea that both methods and ethics are strengthened conceptually and practically when researchers impose the characteristics and functions of each concept onto the other. The exploration begins with a deceptively simple question: How do we “do” ethics? Although ethics is often considered a philosophical stance that precedes and grounds action, it is a value-rationality that is actually produced, reinforced, or resisted through practice. Very quickly, indeed immediately, ethics, when practiced, becomes a matter of method. Likewise, as we act, our every choice has ethical consequences. We might think deliberately about this when we “shop local” or help an elderly person across the street, but in contexts of research and technology design, especially as imagined by the current administrative and regulatory bodies, it can often skip our notice. Yet, our decisions and subsequent actions—large and small—produce and use a particular ethic.

It can be argued that considering how ethics are methods and vice versa is inevitable at this particular moment in time. We are in the midst of an era when the axiological commandment of “doing the right thing” is not at all straightforward. Big data, artificial intelligence, and computational forms of reasoning and analysis have wide-reaching implications for how we act and what positions we take toward our technologies in the course of our practices of inquiry. Twenty years ago, we might have asked, “Should we define and treat public blog texts as intellectual property of humans, or human subjects in and of themselves?” Since then, as storage capacity, network size, computational capacities, and processing speeds grow, the questions seem more and more complicated: Which parts of the computational processes function

as agents or actors, whose decisions should be assessed in how ethical they are? Do we assign responsibility and accountability to the persons who designed the terms of service agreement for the Facebook quiz to allow third parties (in this case, Cambridge Analytica) to access the data of the friends of the quiztaker, or to the API that defines what data should be collected, or to the software that correlated the first data set with other data points held by various data brokers to create a detailed psychometric profile of the users? In the recent case of Cambridge Analytica, the data passed through countless hands, enabling “Cambridge Analytica to connect these psychometric Facebook profiles to actual voters and offer their clients the ability to tailor advertisements to detailed psychometric profiles” (Metcalf & Fiesler, 2018).

We are among those who believe we have only touched the tip of the iceberg in terms of comprehending the impact of the technological at the epistemological and ontological levels, particularly as this influences how we practice (social) science (Floridi, 2012; Kitchin, 2014). As Jenny Davis (2017) notes, methodological innovation no longer seems to mean attempting to “answer particular questions better” which traditionally is a question of “validity” that is contingent on particular methodological techniques; rather, it has become about “asking questions we didn’t know we had” (np). Farida Vis (2013) emphasizes that this is not only a matter of reflecting on our own roles as researchers but also casting “a critical eye on the tools we use” and moving “beyond discussions about their perceived ‘black box’ nature” (np).

Of course, we have seen signs of paradigmatic shifts for some time now. The advent of the Internet refocused researchers’ attention on particular regulatory concepts like “privacy” and “informed consent”. The capacities of the Internet facilitate ways of being and forms of information production and flow that challenge basic definitions around data protection, what it means to be human or a human subject, in ethics regulation terms. Even 10 years ago, the idea that “data are people,” which is expressed as a basic premise in the widely read article by Metcalf, Keller, and boyd (2016), would be highly contested, if not labeled unthinkable. After more than 25 continuous years of monumental technological advancement, it may be difficult to see just how much our everyday epistemic conceptualizations have altered.

This Special Issue contributes specific cases and fine-grained conceptual distinctions to ongoing discussions and critiques of data-driven research models and big data analytics (cf. Beer, 2017; boyd & Crawford, 2012; Cheney-Lippold, 2017; Leurs, 2017; Zook et al., 2017). Seeking inspiration from a variety of theoretical paradigms (i.e., critical theory, [new] materialism, feminist ethics, theory of cultural techniques) and frameworks (i.e., contextual integrity, deflationary perspective, ethics of care), the contributors problematize the “assumptions and trends in big data research” (Luka & Millette, this issue) and point out the “crisis in accountability”

that emerges from uses of digital data to make societal interventions (Pink & Lanzeni, this issue). Our collaborators acknowledge the limitations of approaches that focus primarily on what algorithms do to people (Magalhães, this issue), but highlight the risk inherent in the affordances of machine learning (McQuillan, this issue). They emphasize the conceptual gaps and administrative simplification that often hamper how researchers operationalize ethics (Whelan, this issue; Zimmer, this issue) and hone in on the particular sensitivity of specific “data points” in big data research (Karppi, this issue; Light, Mitchell, & Wikstrom, this issue; Massanari, this issue). Reading the contributions as a whole, we believe they argue that treating ethics as methods usefully compels researchers to focus on the basic idea that ethics matter when they are alive, or enacted. That is to say, ethics that matter are not applied, but produced. It is within our choices at various junctures that our decisions create actions, which in turn have consequence. Here, the moral compass directs action or movement, meaning that the axiological is always being developed, produced, or reproduced alongside the methodological.

Overall, this collection of essays aligns with Metcalf et al.'s (2016) observation that there is no easy consensus on whether or not “big data research methods should be excluded from or forced to meet existing norms, whether existing norms should be made to accommodate the special circumstances of big data, or whether entirely new norms and institutional commitments are needed” (p. 2). The scholars involved in this Special Issue ask and offer a range of answers to the question of how to configure ethics through a methodological framework in the context of the prevalence of big data, neural networks, and automated, algorithmic governance of much of human socia(bi)lity.

Ethics, Methods, and Accountability

As Markham (this issue) summarizes in the afterword of this Special Issue, it is impossible to standardize or universalize what constitutes the ethically correct actions in technology design and research contexts, not least because we cannot predict what will happen as a result of our choices. Obviously, a belief in and understanding of core human rights, plus basic human decency, plus common sense, can help us make good decisions to minimize the potential harmful impact of our actions to persons or society. But our choices are never simply a matter of using our personal calculus: regulations, laws, and policies governing research also define ethics for us. They do so in necessarily generalized terms to provide broad scope and applicability. More informal structural norms we encounter in our professional practice may prescribe particular additional guidelines. So even though ethical decision making might be ideally grounded and produced in practice, through iterative judgments in specific cases, they are presently locked in the structure side of the structure/agency, or structuration cycle (Giddens, 1979). Or, to employ a Foucauldian framing, ethical guidelines become hardened

into incorrigible and obdurate governmentalities (Dean, 1999; Foucault, 1977-1978/2007). This is not an easy situation to unpack, since it is associated with over-regulation in some research-related domains and under-regulation in others.

Meanwhile, the cases of dubious or thoughtless ethical behavior multiply, creating a worrisome ethics crisis. Major elements in this tangle include an “anything-goes” startup attitude that pushes technology designs into the market without adequate scrutiny for their potential impacts, glacially slow updates to conceptual and regulatory guidelines, a push for ever-more precise profiling of people at any cost, and a continued rapid pace of technological transformations that continually change the rules of the game (Markham 2015, Markham and Buchanan 2015).

Alongside this broad crisis, we can identify an equally important failure in the academy. Many, if not all, systems of research and ethics governance do not facilitate and support social research that needs to be done, because it has been traditionally been prohibited. For example, research on children is woefully sparse because children are improperly classified as a single standardized category of vulnerable subjects. As another example, corporate researchers, academics, journalists, and engineers all have different ethical norms governing and guiding their research even as their domains of study overlap. Our understanding of what ethics means and how it might be best enacted therefore suffers. It is our position that embracing a methodological framework of ethics enables productive discussions across domains.

At the level of practice, defining ethics as method draws attention to the epistemic rather than the ontological, or how ethics are enacted, rather than what they are. If taken seriously, this stance brings the concept of ethics from an easily reified abstraction to an always-already emergent and deeply contextual personal practice: anyone doing inquiry of any sort can reflect on how they are personally performing the means. This becomes a necessarily case-by-case process. As Markham writes in 2006, “reflexively interrogating one’s methods of inquiry shifts attention away from codes of conduct imposed from the outside and reveals hidden ethical practices from the inside” (Markham, 2006, p. 39). It also may be used to reflect on how ethics are built not just through the paths that are created when we make a decision to go one way or another at a critical juncture but also in the paths that are closed off, unfollowed, or neglected.

In broad strokes, this approach blends situational ethics and a feminist ethic of care in ways that enact what Judith Simon (2015) discusses as “distributed epistemic responsibility.” As a paradigm, situational ethics builds on the work of the Episcopal priest Joseph Fletcher. Fletcher wrote in 1966 that all choices should be based on the context and circumstance of their particular situation, and not on some universal law. He listed “Love” as the only exception to this situationalism—or rather than an exception, a key mind-set that fosters “good” choices. Although matters of the heart may appear incidentally in contemporary discussions of research ethics, they constitute a

core feature of Markham's (2006) ideas about how ethics are, or should be, enacted as method. "Heart," she explains, is "an amalgam of consciousness, mindfulness, honesty, and sensitivity," and an ethical researcher is one who "works from the center," which entails "being knowledgeable and prepared; present and aware; adaptive and context sensitive; and honest or mindful" (p. 44).

This stance is a hallmark of a feminist ethic of care, yet as defined by Markham here, it seems less about love, specifically, and more about readiness, which draws on Allport's (1935) definition of the sort of preparation or attitude that is essential to make "satisfactory observation, pass suitable judgment, or make any but the most primitive type of reflex response" (p. 806). In the entangled contexts of data-driven research, however, the consciousness and adaptability that come from readiness may seem ineffective when the individual researcher might play only a minor role—working in large teams, encountering only whatever has been scraped and filtered by an API, or studying data long removed from its human origins. We argue that despite the challenges posed by automated systems, returning to a baseline definition of method as a series of axiological choices can help us move toward more appropriate understandings of how epistemological responsibility (Simon, 2015) moves from being assigned (or distributed) to being enacted, in that we turn our focus toward the specificities of the actions that constitute choices that matter.

Methods as Ethics: How the Tools Under Our Methods Produce Ethics

The surface level accuracy of big data analytics implicitly valorizes and explicitly fosters an old school correspondence theory of truth. The concept of "data" presumes the possibility that human behavior can be traced, isolated, collected as data points, and measured. Big data, as a process of aggregating and performing calculations on diverse and huge sets of data points at high speed, yields promising "truths" about who we are and what we therefore want. The tools used to yield these findings are at some level making interpretations about what the data mean, but the interpretive stages occur early in the processes of testing so that by the time the data analysis yields an effective result, it is highly rule based and computational. Thus, even as we see a renaissance of nuanced interpretivism in the study of humans and society, the discourses around data science promote, whether deliberately or not, a return to the most conservative sorts of positivism.

Along with the authors of this Special Issue, we define and confront this as an emerging grand narrative. The basic premise is that knowledge derived from data analytics is true (or has strong truth value) because of the objective qualities of data, their means of collection and analysis, and the sheer size of the data set. This deceptive premise has been repeated in many forms: in the late 1990s, distance education was heralded as a way to "deliver" knowledge, as if the essence of knowledge was embedded in the content of the material

transmitted online to the learner. Now, platform and service providers assert that their algorithms are impartial, functioning through rule-based calculations on objective data points without any "subjective" interference. This standardization supposedly secures their position as "legitimate brokers of relevant knowledge" (Gillespie, 2012, p. 180). The rule-based process—or as we emphasize, the method—of the algorithm indeed yields almost divine-quality results, especially when we see these outcomes as eerily-appropriate advertising on our own screens. The by-product of common discourse around big data not only diminishes or removes the qualitative aspects of behavior and experience that form the data in the first place but also removes the human from the process of analysis. This situates data science as a process of analysis performed by the tool, which removes decision making and judgments from the endeavor.

Historically, the early 19th century science of craniometry is a good example of how our tools both use and produce an ethic, or ethos, which carries significant axiological weight in the design and conduct of scientific inquiry. In the scientific tradition of craniometry, skull size was associated with intelligence, and it just so happened that Caucasians had larger skulls than people categorized into other races. The choice to define intelligence as a solid and measurable variable can be seen as a methods move that, over time and restatement, creates plausibility. We have rejected measuring skull size as a valid method of determining intelligence, but the ethos of reducing intelligence to a numeric value (e.g., bell curves, IQ test scores) remains. This can be seen as an ethics move, in that the use and valuation of particular methods produce or reinforce particular ways of knowing.

The authors in this Special Issue draw out contemporary corollaries. Relevancy, for example, is used in big data analytics as if it is not a subjective category but a natural choice for how search engines should yield results for users. But as computer scientist Joanna Bryson and her colleagues found, algorithms draw on the same biases that humans have (Caliskan, Bryson, & Narayanan, 2017). In this and other cases, the ethos, it turns out, disenfranchises already-marginalized groups (cf. Eubanks, 2018; Noble, 2018). Therefore, algorithms enact a different form of an-ism. As Karppi (this issue) points out algorithms that use relevancy in predictive policing make distinctions and mark "hot spots," which shapes how people are treated by authority figures. This in turn can influence the resources certain people have access to, how likely they are to succeed at various endeavors, or how likely they are to be penalized by governmental agencies (see also Brayne, 2017).

Often, the ethic is engendered not directly through the actions of the researcher, but indirectly through the absence of questioning the validity of variables in a world that has long since discovered (Kuhn, 1962) that our basic paradigms about what things are, or how they work, are not naturally "true," but an outcome of debate, persuasion, and other social interactions among scientists. If we apply some of this critical reflexivity toward tools such as APIs, which extract

certain information as data while ignoring other information as non-relevant, we can begin to move beyond the obfuscating critique that they operate as black boxes, and begin to question what they are actually producing.

As Rob Kitchin (2014) notes, social research methods have traditionally had to “extract insights from scarce, static, clean and poorly relational data sets,” whereas in the era of “data-driven” research, the challenge is to cope “with abundance, exhaustivity and variety, timeliness and dynamism, messiness and uncertainty, high relationality, and the fact that much of what is generated has no specific question in mind or is a by-product of another activity” (p. 2). We use machine learning systems to cope with this messy abundance, since the size of any single data set and the aggregated complexity of multiple data sets are beyond the processing capacity of human cognition. However, machine learning is becoming “a methodological substrate for knowledge and action . . . a kind of dark matter that invisibly distorts the distribution of benefits and harm” (McQuillan, *In Press*, p. XX). And, while it is, by now, a well-known and perhaps battered point that datafication has transformed how social research is conducted, more subtly, it impacts what we think we are doing when we conduct said research. A dangerous erosion of the role and meaning of interpretation seems to accompany the shifts that “data-driven” research and interventions have introduced (Markham, 2017).

While this epistemic shift toward “algorithmic knowledge production” (Metcalf et al., 2016, p. 6) highlights the differences and tensions between what we have conventionally considered good social research and what we think big data research is, efforts are being made to synergize the overlaps between the two. Pink and Lanzeni (this issue) suggest that future-focused anthropology and big data research have some things in common. Both share improvisational data gathering methods, encounter the world as it unfolds, focus on the futures, and have interventional ambitions. Somewhat similarly, Halford and Savage (2017) suggest “symphonic social science” as a way forward. Symphonic social science can be achieved through compiling multiple different data sources into a single, unified, and often repeated “refrain,” which in turn makes the arguments created via such methods more persuasive (Halford & Savage, 2017). The similarities (multiple data sources, emphasis on correlation, and use of visualization), and differences (i.e., symphonic SR includes rich theoretical awareness, carefully chosen data, focus on long-term trends) between symphonic social science and big data research are what Halford and Savage (2017) think could extend both and offer a methodological solution to the epistemic shift underfoot.

Agency, Responsibility, and Morality

If we agree that choices (in research) should be contextual and situational and that moral situations involve a relationship between, at least, an originator of an action—or a moral agent and the recipient of actions—or a moral patient, then

the actions of the agent and the harms or benefits to the patient can and should be evaluated (Floridi & Sanders, 2004). This formula seems reasonable: It implies that the ethicality of any situation can be assessed either from the perspective of the agent (i.e., responsibility based evaluation) or from the perspective of the patient (i.e., rights based evaluation). But it offers a too-simple binary in the context of datafication, digitalization, and automation.

Perceived agency for ethical decision making has become complicated in the context of machine learning. Big data, which Kitchin (2014) describes as voluminous, swift, diverse, exhaustive, fine grained, relational, and flexible, comes with a “class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society” (Matthias, 2004, p. 177). How do we attribute moral agency, moral patiency, or shared responsibility when neural networks are “obfuscated by nature” (McQuillan, this issue), or when algorithms function as cultural techniques that seem to have their own agency (Karppi, this issue)? At the same time, should we accept the idea that we no longer have agency? Magalhães (this issue) contends that despite our inability to comprehend algorithms, human actors, even non-expert end users, are still capable of perceiving what algorithms do—an insight that grants people some moral power and agency.

We can call this complication of locating moral agency and responsibility a wicked problem. There are no straightforward boundaries, definitions, or answers. Rather, there are only questions to be continually addressed. As Gunkel (2017) notes, almost drily

how we decide to respond to the . . . machine question will have a profound effect on the way we conceptualize our place in the world, who we decide to include in the community of moral subjects, and what we exclude from such consideration and why. (p. 245)

The dilemmas are not new, of course. Discussions of machine ethics often reference the five laws of robotics that the science fiction writer Isaac Asimov posited already in the early 20th century. The first industrial revolution is peppered with machines that critics denounced for dehumanizing worker and separating the mind from the hand, conception of the purpose of work from its execution by labor (Braverman, 1974). Shelley’s *Frankenstein* yields a different sort of speculation about non-human morality. Yet, by and large, human exceptionalism still stands as the dominant ethical paradigm, and it positions technology as a tool created and used by human beings. This presumption has significant limitations in the light of recent technological innovations (Gunkel, 2017). Google’s go-, shogi-, and chess playing AlphaZero, their facial recognition deep learning network FaceNet, Microsoft’s hate-spewing chatbot Tay.ai were all designed to create their own instructions and evolve their behavior. The people who created them thus have limited control over what the systems do. If neural and deep learning networks develop

“minds” of their own, but we view ethics as an issue of agency, then what happens to our ethics when agency blurs? At what point can we hold an API or an algorithm accountable for its decisions and actions? Or how can we hope to control them (cf. Karppi, this issue; McQuillan, this issue)?

While there is no consensus on whether we can assign moral agency or moral patiency to machines and systems, “what is not debated is the fact that the rules of the game have changed significantly” (Gunkel, 2017, p. 237). Even if we do not think that machines can or should have a moral status, it seems that the ability to act has escaped human confines. The time to ask ourselves whether it is ethical to build systems that have the ability to act without having the ability to take on moral agency, too, has passed, as we have already built them.

Returning to the question of how we can identify morality in these entanglements of agency, we agree with Silverstone (2006) that it is far more important to consider matters of responsibility and accountability. These concepts call our attention to the idea that someone needs to take responsibility for responding, when something goes awry. This shifts the focus in a pragmatic and future-oriented way.

The Role of the Researcher

If “unraveling the intricate tapestry of method and ethic . . . involves partitioning what appears to be a smooth flow of one’s choices and movements during the entire research project” (Markham, 2006, p. 39) then how do we accomplish that in the context where automated systems play an increasing role in those movements, and possibly alter what we experience as choices? Most of us can probably agree that good researchers strive to find situated, context-appropriate ways for linking their habitual decision making with morality. But there is less consensus on how to actually accomplish this. We suggest it is a matter of interrogating what is happening along the string of actions that eventually lead to the construction of data, focusing on human practices and nonhuman processes that function interpretively, and reflecting on various levels and stages of possible impact of such actions, constructions, and interpretations.

Together, the collaborators of this Special Issue coalesce around two broad suggestions for building more reflexive practice:

1. Develop or employ a heuristic of ethical decision making, that is (a) both more practical and less abstract than most ethics guidelines and can (b) help you see the research situations you are immersed and invested in in a new light.

For example, Zimmer (this issue) offers an ethics assessment heuristic based on Nissenbaum’s (2004, 2010) “privacy as contextual integrity” framework. He points out that this heuristic can empower researchers to be “more attentive of the normative bounds of how information flows within

specific contexts” (Zimmer, in Press). To demonstrate his chosen heuristic in action, Zimmer (this issue) applies its nine steps to the much-critiqued 2016 case, where a Danish graduate student released a large amount of OKCupid user data to the public with the claim that it was “for science” and OK, because “the data was already public.” Zimmer persuasively shows how much harder it would have been to dismiss the moral and ethical problems of the “already public” claims, if the data scrapers and sharers would have assessed its contextual integrity and not just its “public” accessibility.

A different heuristic is offered by Luka and Millette (this issue). They describe speculation as “a gathering of possibilities” that allows researchers to practice care by making their own voice one among many, becoming aware of, and sensitive to, their own positions, biases, and power. They queer Hannah Arendt’s (1961) concept of action¹ to analyze the “unacknowledged costs involved in data production and analysis.” This might be enacted through processes of superseding, supplanting, or augmenting big data by small, lively, or thick data (or the other way around) within the same project (cf. also Latzko-Toth, Bonneau, & Millette, 2017). Whereas Zimmer’s approach relies on the framework of “contextual integrity” to be able to take on an alternative perspective in regard to the research project, Luka and Millette suggest designing studies with, rather than upon others, to create a context-specific heuristic:

2. Seek inspiration from colleagues who deal with sensitive topics, high-risk research situations, and/or vulnerable populations. They’re developing tools and skills that can come in handy in the context of the epistemic and axiological shifts, where we are all potentially vulnerable, and all data are potentially sensitive (cf. Tiidenberg, 2018).

Sharing experiences from their own sensitive research project with men who have sex with men, Light et al. (this issue) show how a particular aspect or a “data point” can become particularly sensitive in specific contexts and situations. They offer an example wherein location becomes such a key factor for big data ethics. Operationalized through stages of gathering, analyzing, and presenting, as well as archiving and deleting data, the authors show how the most ethical choices may actually be those that are methodologically counterintuitive. They emphasize that not all that can be collected should be; that data visualization has its own politics, and thus ample ethical problems; and that even though storing data and sharing code is lauded in social research as a tool of increased validity and reliability, an ideologically communal act, or a way to get more “bang for public buck,” the most ethical choice may be to delete both data and scripts.

Sharing her insights from studying extremism—arguably the current plague worldwide, and an issue undoubtedly entangled with big data, algorithmic sociality, and social media—Massanari (this issue) highlights the need to mitigate

researcher risk under these new conditions. However, as she emphasizes, the need to reduce harassment and intimidation of academics clashes with the dominant norm of academic visibility and microcelebrity. From this junction, Massanari provides concrete suggestions for professional practices that work well to support researchers, share best practices, and build stronger and broader understandings of risk.

These two broad suggestions may resonate with our own attitudes and inspire us. Yet it is also quite clear that they do not easily link with or accomplish formal requirements of funding agencies, PhD programs, and ethics oversight institutions. This can create some cognitive dissonance for researchers seeking to do what is actually ethical, when it may not align with extant regulatory guidelines (Whelan, this issue). This dissonance is likely quite common at present, as it is well understood that “insofar as research ethics regulations and conceptual frameworks are responding to the conditions of knowledge production that precede the epistemic shift towards algorithmic knowledge, (big data) research and the extant research ethics regimes will be mismatched” (Metcalf et al., 2016, p. 6).

There are new models to help ease such dissonance and link methods to ethics and vice versa. Two influential groups are worth mentioning for their steady insistence that as scholarly communities, we owe it to ourselves, future generations of scholars and humans, and perhaps, idealistically, the state of the world, to push against administrative, corporate, regulatory, and legislative frameworks that fail to accommodate the needs of contemporary ethical practice in digital media use and social research: The ethics working group of the Association of Internet Researchers (AoIR)², who are now working on their third iteration of international guidelines for ethical decision making in social research, and American University’s Center for Media and Social Impact³, who have for many years developed industry-specific codes of best practice that challenge outdated interpretations of copyright.

Our special issue joins this growing number of voices imagining, experimenting with, and sharing the practices of enacting ethics as methods and methods as ethics. The authors all agree in different ways that we cannot afford to be paralyzed by the wicked problems and the uncomfortable researcher subjectivities prompted by technological, political, economic, and institutional complexities. We hope the contributions in this Special Issue inspire readers to challenge how ethics are conceptualized within ill-fitting regulatory and legal frameworks, explore how our tools and techniques carry and produce an ethos that over time becomes taken for granted, and operationalize epistemological concepts that fit the complexities of agency and accountability in data-entangled research contexts.

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) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. as radically free, potentially rationality-preceding, aspirational and speculative activity
2. <http://aoir.org/ethics/>
3. <http://cmsimpact.org/program/fair-use/>

ORCID iD

Annette N Markham  <https://orcid.org/0000-0001-8152-2473>

References

- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.
- Arendt, H. (1961). “What is freedom.” *Between past and future, six exercises in political thought*. New York, NY: The Viking Press.
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13. doi:10.1080/1369118X.2016.1216147
- boyd, d., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679. Retrieved from <http://doi.org/10.1080/1369118X.2012.678878>
- Braverman, H. (1974). *Labour and monopoly capital: The degradation of work in the twentieth century*. New York, NY: Monthly Review Press.
- Brayne, S. (2017). Big Data Surveillance: The Case of Policing. *American Sociological Review*, 82(5), 977–1008. Retrieved from <http://doi.org/10.1177/0003122417725865>
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356, 183–186.
- Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. New York: New York University Press.
- Davis, J. (2017). *Big Data and the epistemological renaissance*. Available from <https://thesocietypages.org/cyborgology/2017/06/05/big-data-and-the-epistemological-renaissance/>
- Dean, M. (1999). *Governmentality: Power and rule in modern society*. London, England: SAGE.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York, NY: St. Martin’s Press.
- Fletcher, J. F. (1966). *Situation ethics: The new morality*. Philadelphia, PA: Westminster.
- Floridi, L. (2012). Big Data and their epistemological challenge. *Philosophy and Technology*, 25, 435–437.
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14, 349–379.
- Foucault, M. (2007). *Security, territory, population: Lectures at the Collège de France 1977–1978* (G. Burchell, Trans. & M. Senellart, Ed.). New York, NY: Palgrave Macmillan. (Original work published 1977–1978)

- Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis*. Los Angeles: University of California Press.
- Gillespie, T. (2012). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies* (pp. 167–195). Cambridge, MA: MIT Press.
- Gunkel, D. J. (2017). Socialbots and the question of ethics. In R. W. Gehl & M. Bakardjieva (Eds.), *Socialbots and their friends: Digital media and the automation of sociality* (pp. 230–249). New York, NY: Routledge.
- Halford, S., & Savage, M. (2017). Speaking sociologically with Big Data: Symphonic social science and the future for Big Data research. *Sociology*, 51, 1132–1148.
- Hayles, K. (2015, February 26–27). *Future anterior, derivative writing, and the cognitive technosphere*. Keynote at Thinking with Algorithms: Cognition and Computation in the Work of N. Katherine Hayles, Durham, NC.
- Karppi, T. (IN PRESS). “The computer said so”: On the ethics, effectiveness, and cultural techniques of predictive policing. *Social Media + Society*.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1, 1–12.
- Kuhn, T. S. (1962). *The structure of scientific revolutions* (1st ed.). Chicago, IL: The University of Chicago Press.
- Latzko-Toth, G., Bonneau, C., & Millette, M. (2017). Small Data, thick data: Thickening strategies for trace-based social media research. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 199–214). Beverly Hills, CA: SAGE.
- Leurs, K. (2017). Feminist data studies: Using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review*, 115, 130–154.
- Light, B., Mitchell, P., & Wikstrom, P. (IN PRESS). Big Data, method and the ethics of location: A case study of a hookup app for men who have sex with men. *Social Media + Society*.
- Luka, M. E., & Millette, M. (IN PRESS). (Re)framing Big Data: Activating situated knowledges and a feminist ethics of care in social media research. *Social Media + Society*.
- Magalhães, J. V. (IN PRESS). Do algorithms shape character? Considering algorithmic ethical subjectivation. *Social Media + Society*.
- Markham, A. N. (2003). Critical junctures and ethical choices in Internet ethnography. In M. Thorseth (Ed.), *Applied ethics in Internet research* (pp. 31–51). Trondheim: NTNU University Press.
- Markham, A. N. (2006). Method as ethic, ethic as method. *Journal of Information Ethics*, 15(2), 37–55.
- Markham, A. N. (2015). Producing ethics [for the digital near future]. In R. A. Lind (Ed.), *Producing theory in a digital world 2.0: The intersection of audiences and production in contemporary theory* (Vol. 2, pp. 247–256). New York, NY: Peter Lang.
- Markham, A. N. (2017). Troubling the concept of data in digital qualitative research. In U. Flick (Ed.), *Handbook of qualitative data collection* (pp. 511–523). London, England: SAGE.
- Markham, A. N. (IN PRESS). Afterword: Ethics as impact: Moving from error-avoidance and concept-driven models to a future-oriented approach. *Social Media + Society*.
- Markham, A. N., & Buchanan, E. (2015). Ethical considerations in digital research contexts. In J. D. Wright (Ed.), *Encyclopedia for social & behavioral sciences* (pp. 606–613). Waltham, MA: Elsevier.
- Massanari, A. (IN PRESS). Rethinking research ethics, power, and the risk of visibility in the era of the “alt-right” gaze. *Social Media + Society*.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6, 175–183.
- McQuillan, D. (IN PRESS). People’s councils for ethical machine learning. *Social Media + Society*.
- Metcalfe, J., & Fiesler, C. (2018, March 18). One way Facebook can stop the next Cambridge Analytica: Give researchers more access to data, not less. *Slate*. Retrieved from <https://slate.com/technology/2018/03/cambridge-analytica-demonstrates-that-facebook-needs-to-give-researchers-more-access.html>
- Metcalfe, J., Keller, E. F., & boyd, d. (2016). *Perspectives on Big Data, ethics, and society*. The Council for Big Data, Ethics, and Society. Retrieved from <http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/> (Accessed November 1, 2017).
- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79, 119–159.
- Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.
- Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: New York University Press.
- Pink, S., & Lanzeni, D. (IN PRESS). Future anthropology ethics and datafication: Temporality and responsibility in research. *Social Media + Society*.
- Silverstone, R. (2006). *Media and morality: On the rise of the Mediapolis*. Cambridge, UK: Polity Press.
- Simon, J. (2015). Distributed epistemic responsibility in a hyper-connected era. In L. Floridi (Ed.), *The Onlife manifesto: Being human in a hyperconnected era* (pp. 145–159). London, England: Springer Open.
- Skeggs, B. (2017, October 30–November 1). *What are the consequences of tracking, trading and sub-priming the subject through stealth?* Keynote at Digital Existence II: PRECARIOUS MEDIA LIFE conference, The Sigtuna Foundation.
- Tiidenberg, K. (2018). Ethics in digital research. In U. Flick (Ed.), *The sage handbook of qualitative data collection* (pp. 466–479). London: SAGE Publications Ltd. doi:10.4135/9781526416070.n30
- Tiidenberg, K. (2018). How not to be an assh*le? Research ethics, vulnerability and trust on the internet. In J. Hunsinger, L. Klasturp, & M. Allen (Eds.), *International handbook of Internet research* (pp. XX–XX). New York: Springer.
- Vis, F. (2013, October). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*. Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/4878/3755>
- Whelan, A. (IN PRESS). Ethics are admin: Australian human research ethics review forms as (un)ethical actors. *Social Media + Society*.
- Zimmer, M. (IN PRESS). Addressing conceptual gaps in Big Data research ethics: An application of contextual integrity. *Social Media + Society*.

Zook, M., Barocas, S., boyd, d., Crawford, K., Keller, E., Gangadharan, S. P., . . . Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS Computational Biology*, 13, e1005399.

Author Biographies

Annette N Markham is a professor MSO of Information Studies at Aarhus University, Denmark, and Affiliate Professor of Digital Ethics in the School of Communication at Loyola University, Chicago. Annette is internationally recognized for developing epistemological frameworks for rethinking ethics and qualitative methods for digitally saturated social contexts. A long-time member of the digital research community, Annette conducts sociological and ethnographic studies of how identity, relationships, and cultural formations are constructed in and influenced by digitally saturated socio-technical contexts. For more information, please see annetmarkham.com.

Andrew Herman received his BA in Government from Georgetown University and his PhD in Sociology from Boston College. American by birth and Canadian by choice, Dr Herman taught at Boston College, Drake University and College of the Holy Cross before joining the Communication Studies department at Laurier in 2004. He has written widely in the field of social theory, media, and culture and his work has appeared in scholarly journals such as *Cultural Studies*, *Critical Studies in Media Communication*, *South Atlantic Quarterly*, and *Anthropological Quarterly*. Among his many publications is his book, *The “Better Angels” of Capitalism: Rhetoric,*

Narrative and Moral Identity Among Men of the American Upper Class (Westview, 1999) and his edited collections, *Mapping the Beat: Popular Music and Contemporary Cultural Theory* (Blackwell, 1997), *The World Wide Web and Contemporary Cultural Theory* (Routledge, 2000). His most recent book is *Theories of the Mobile Internet: Materialities and Imaginaries* (Routledge, 2015). He is currently working on two research and writing projects: “Cats that look like Kittler”: Internet Cats and the Medium Materialities of the World Wide Web, 1995-2015 and New Spirits of Informational Capital(ism): Cultures of Innovation in Canadian Tech Clusters and Entrepreneurial Ecosystems.

Katrin Tiidenberg (PhD) is an associate professor of Social Media and Visual Culture at the Baltic Film, Media, Arts and Communication School of Tallinn University, Estonia and a post-doctoral researcher at Aarhus University, Denmark. She is the author of “*Selfies, Why We Love (And Hate) Them*” (2018), as well as “*Ihu ja hingega internetis, kuidas mõista sotsiaalmeediat*” (*Body and Soul on the Internet—Making Sense of Social Media*) (in Estonian, 2017). Katrin is a long-time member of the Association of Internet Researcher’s Ethics Committee, a founding member of the Estonian Young Academy of Sciences, second time board member of the Estonian Sociology Association. She is currently writing and publishing on selfie culture, digital research ethics, and visual research methods. Her research interests include visual culture, sexuality, and normative ideologies as mediated through social media practices and networked technologies. For more information, please see kkatot.tumblr.com