

“You Social Scientists Love Mind Games”: Experimenting in the “divide” between data science and critical algorithm studies

Big Data & Society
January–June 2019: 1–11
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951719833404
journals.sagepub.com/home/bds



David Moats¹ and Nick Seaver²

Abstract

In recent years, many qualitative sociologists, anthropologists, and social theorists have critiqued the use of algorithms and other automated processes involved in data science on both epistemological and political grounds. Yet, it has proven difficult to bring these important insights into the practice of data science itself. We suggest that part of this problem has to do with under-examined or unacknowledged assumptions about the relationship between the two fields—ideas about how data science and its critics can and should relate. Inspired by recent work in Science and Technology Studies on interventions, we attempted to stage an encounter in which practicing data scientists were asked to analyze a corpus of critical social science literature about their work, using tools of textual analysis such as co-word and topic modelling. The idea was to provoke discussion both about the content of these texts and the possible limits of such analyses. In this commentary, we reflect on the planning stages of the experiment and how responses to the exercise, from both data scientists and qualitative social scientists, revealed some of the tensions and interactions between the normative positions of the different fields. We argue for further studies which can help us understand what these interdisciplinary tensions turn on—which do not paper over them but also do not take them as given.

Keywords

Algorithms, data science, intervention, reflexivity, interdisciplinarity, Science and Technology Studies

This article is a part of special theme on Algorithmic Normativities. To see a full list of all articles in this special theme, please click here: https://journals.sagepub.com/page/bds/collections/algorithmic_normativities.

Asleep at the wheel?

In November 2017, the *New York Times* published an opinion piece titled “The Ivory Tower Can’t Keep Ignoring Tech,” by Cathy O’Neil, a data scientist and popular critic of Big Data. In her piece, as in her book, *Weapons of Math Destruction* (2016), O’Neil made an argument familiar to readers of this journal: algorithmic systems have accrued immense power, processing ever more data in ever more domains, and they exert this power in relative obscurity, hidden from the prying eyes of critics and the people whose lives they affect.¹ It is crucial, she argued, to subject these systems to outside examination if we want to mitigate the growing variety of algorithmic harms.

As O’Neil’s piece came out, we were on our way to the third meeting of the Algorithm Studies Network outside of Stockholm to workshop an early version of the piece you are reading now. This network brought together an international group of critical and interpretivist scholars from the social sciences and humanities to share work on the social lives of algorithmic systems. By 2017, the network was only one among many

¹TEMA T, Linköping University, Linköping, Sweden

²Department of Anthropology, Tufts University, Medford, USA

Corresponding author:

David Moats, TEMA T, Linköping University, Linköping 581 83, Sweden.
Email: david.moats@liu.se



organizations, events, and publications dedicated to the social study of algorithms. As early as 2015, this field was so large and dispersed that one of us curated a reading list on “critical algorithm studies” (<https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>) in an effort to draw together critical work on algorithms from the previous decades.

Like many of our colleagues, we were surprised by the turn O’Neil’s argument took: “Academics have been asleep at the wheel,” she wrote; “There is essentially no distinct field of academic study that takes seriously the responsibility of understanding and critiquing the role of technology—and specifically, the algorithms that are responsible for so many decisions—in our lives” (O’Neil, 2017). At every level, this seemed baffling. Algorithms were such a hot topic in the critical study of technology that their trendiness was a topic of academic jokes. The field of Science and Technology Studies (STS) had existed for decades, largely founded on the goal of “understanding and critiquing the role of technology in our lives.” What did it mean to suggest that this topic and field did not exist?

As ethnographers with experience studying programmers and data scientists (Seaver) and working with them to develop research tools (Moats), we recognized O’Neil’s argument as a familiar kind of boundary work (Gieryn, 1999). Intentionally or not, O’Neil had suggested that “real” critical work on algorithms—the kind that might make a political difference and claim the epistemic high ground—would come from data scientists themselves, not from outside researchers with questionable authority on the topic. Her call for “robust research” which “pushes against the most obvious statistical, ethical, or constitutional failures” also seemed to sideline research which raised more fundamental questions about the enterprise of algorithmic knowledge or modes of decision making, while relying on unspoken normative assumptions about what constituted “robustness” or even “research” in the first place.

In our own work and that of many of our colleagues, this collision of normativities was a defining experience: our interlocutors, trained in academic computer science programs and holding “technical” positions as engineers and scientists were rarely obliged to read or understand the critical, interpretivist work written about them. And while our work was animated by a desire to understand their field, our disciplinary commitments meant that we could not fully embrace their ideals of research or knowledge. At times it felt like we were speaking different languages.

Frustrated by familiar cycles of overreach and outrage, we wondered how we could intervene in this dynamic. What would happen if we encouraged people like our interlocutors to engage with critical work but to

do so on their own terms? This paper is about the first stages of an experiment in which we would provide data scientists with a collection of texts representing work in “critical algorithm studies” and ask them to analyze it using their own disciplinary tools—using algorithms to make sense of critical algorithm studies. What sense would they draw, coming to this work with their own normative assumptions and techniques and what would they make of our normative positions?

In this paper, we will discuss the play of normativities during the project’s planning stage, in a series of encounters with both computer scientists and STS scholars. These encounters demonstrate how different normativities (on both sides) shaped understandings of the presumed difference between the two fields. In trying to intervene in this social dynamic, that dynamic intervened on us.

Intervening in algorithmic practice

In recent years, a growing body of work by qualitative sociologists, anthropologists, and social theorists has critiqued the turn toward the so-called “Big Data” and algorithms, and the field now called “data science” (Iliadis and Russo, 2016). These researchers have raised concerns about methods of data collection, which smuggle in assumptions about what is real and what matters (Gitelman, 2013); they have criticized quantitative modes of analysis which reduce lived experience to stark categories and metrics (Adams, 2016; boyd and Crawford, 2012), leading researchers to ask reductive questions (Marres and Weltevrede, 2013; Uprichard, 2013); they have shown how systems built on algorithmic logics format the world in biased or unexpected ways (Noble, 2018; Ruppert et al., 2015) while their machinery remains “black boxed,” and inaccessible to public scrutiny (Burrell, 2016; Pasquale, 2015). In short, these researchers have argued that despite its many claims to objectivity and legitimacy, data science is politics by other means.

Yet, it has proven challenging to bring these critiques, which we will call collectively “critical algorithm studies,” into the craft of data science itself. Calls for an “ethical data science” (boyd and Crawford, 2012; Kitchin, 2014) have found some traction in practitioner communities (e.g. Schutt and O’Neil, 2013), and a growing community of technical researchers has pursued projects concerned with bias, fairness, and accountability (e.g. FATML: <https://www.fatml.org/>). Yet, as ethical critiques are taken up, broader epistemological, ontological, and political questions about data science tools are often sidelined. To be clear, computer scientists frequently raise methodological concerns about their discipline (e.g. Lipton and Steinhardt, 2018) but they tend to do so within their own normative

frame, rather than addressing for example the entanglement of politics and knowledge raised by the critical literature. These researchers are also not adverse to bring social science concerns into their work (Lazer et al., 2009; Wallach, 2015) but this also involves a particular version of social science which is already concerned with explanatory claims and quantitative techniques.

Part of the problem, we contend, is that many existing critiques, and efforts to bring critique into practice, take the relationship between data science and its critics for granted. They often rehearse a set of common-sense distinctions between qualitative and quantitative methods, between ethical critics and unethical practitioners, positivist programmers and interpretive ethnographers, and so on. We have both argued against such distinctions in our home fields. Seaver (2015) has described how, in anthropology, ethnographic sensibilities have long been defined in the negative image of mathematical formalism and consequently, contemporary “discoveries” of the complementarity of data science and ethnography are actually just identifying definitional principles that have shaped ethnographic practice since the days of Malinowski. Moats (2017) has questioned why, when ethnographers and qualitative social scientists shadow statisticians and programmers, the tensions and fissures between these two ways of knowing are largely accepted as a natural state of affairs: a tension between “stories” and “numbers.” Indeed, when critical researchers do attempt to collaborate with data scientists, they often realize that their counterparts are well aware of many questions around complexity, politics, and performative effects, but make sense of them in distinctive ways as they make compromises in producing data-driven products (Neff et al., 2017).

Yet, while these disciplinary divisions should not be taken as given, this is not to suppose that they are completely fictitious, as the fallout over O’Neil’s article makes clear. As more researchers work toward ethical, hybrid practices that recognize the assumptions, omissions, and performative effects of algorithmic systems, it is important to better understand the relationships between these fields. Otherwise we risk just critiquing data science as some pallid form of ethnography, a “thin description” or “distant reading,” ignoring the epistemic commitments that many data scientists hold: the point of their work is to reduce, generalize, and categorize. Drawing on our STS sensibilities, we might say that these divergent commitments are not a matter of fundamental, essential differences between methods, disciplines, or social groups, but are rather the results of situated practices and mundane interactions. When we speak of a “divide,” we are not arguing that it is desirable, natural, or inevitable, but

are rather pointing to an empirical phenomenon which manifests in practice as conversational tension, miscommunication, and, sometimes, disputes.

It should be noted that, across STS and allied fields, many scholars have produced frameworks for conceptualizing or managing similar interdisciplinary tensions in practice. Hackathons (Irani, 2015) and Data Sprints (Munk et al., 2019), both types of collaborative, project-oriented workshops, have provided settings for programmers or data scientists and non-technical “topic experts” to gather around shared problems. But as has often been noted, the horizon of possibilities in these interactions is often set by the more technically capable participants (Ruppert et al., 2015) rather than the “qualitatively” oriented ones. In other words, normative stances regarding what counts as “useful” or “interesting” in these collaborations often come from programmers. Various uses of the term “co-laboratory” or “collaboratory” have provided new ways for ethnographers to think through their relationship to their field site but it has rarely resulted in ethnographers doing anything other than observing from the sidelines (Rabinow et al., 2008). In anthropology, George Marcus has proposed the construction of ethnographic “para-sites” (2000), situations where ethnographers and their interlocutors could meet outside of the naturalistic imaginary of the field. The para-site is “a site of alternativity, in which anything, or at least something different, could happen” (2000: 8). While these frameworks are helpful ways of thinking through interdisciplinary encounters, we argue that they rarely put the underlying normative commitments of these disciplines at risk, nor do they offer guidance in cases of non-recognition or situations where collaborations are hard to get started in the first place.

In our effort to explore tensions between the normative commitments of data scientists and qualitative researchers, we drew inspiration from recent efforts in STS to move beyond detached description toward “situated interventions” (Zuiderent-Jerak, 2015) or “making and doing” (Downey and Zuiderent-Jerak, 2016). For example, Zuiderent-Jerak, as a participant-observer in healthcare settings, has used common technologies of his informants like flowcharts and budgets to tease out the normative commitments at play in the field (including his own). These interventionist research programs are interesting because they involve social scientists getting their hands dirty. They are experimental, but not like *in vitro* experiments in a laboratory (Callon et al., 2007) or randomized controlled trials (Adams, 2016). Rather, they are open-ended transactions with social worlds, the results of which cannot be determined in advance. They are thus “experimental” in the sense used by the composer John Cage (1961): “an action the outcome of which is not foreseen” (69).²

While particular interventions may have concrete aims, they are also opportunities to learn about social settings and how actors within them respond.³ Some interventions are modeled on practices of prototyping and testing, but without the connotation of necessary streamlining and improvement. Rather than embodying the teleological vision of a pebble rolling down a hill being progressively smoothed and refined through constant bashing, these experiments seem more like a snowball, collecting more and more debris, breaking into bits and changing shape through contact with the world. These interventions are careful, but messy, and in contrast to classic models of scientific research, they put our own roles and identities at risk through the interaction (Stengers, 2000). As a result, interventions may teach us just as much about our own baggage and assumptions as we do about those of the people we set out to study, as we will explain below.

Encounter one: “It feels like a snake eating its own tail”

From our ethnographic research, it was clear that, while data scientists were familiar with broad-strokes critiques of their work, they rarely read academic critiques of the sort published in journals like *Big Data & Society*. This was not surprising: most of these pieces are written for social science and humanities readers, rather than for programmers or computer scientists. Yet, the implicit (and sometimes explicit) premise of much critical literature seems to be that, if people in power would only listen, then change might happen. Hence, we wanted to confront our interlocutors with this writing—not to convince them of its correctness, but to see what might come of actual engagement with it. Perhaps, we hoped, this might help us (and them) think about how data scientists conceptualize their nascent discipline in relation to its critical others? Would they fall into standard tropes of boundary work? Would they perform a sort of “counter-reading” (Hall et al., 1980)? And most importantly, how could we deliver these critiques and observations in a way which was interesting for them, without jeopardizing our carefully crafted roles as friendly observers or collaborators?

We settled on a simple exercise that could be completed in a workshop setting, hackathon, or even remotely at home. We would collect a corpus of texts (with dates, titles, and abstracts) representing the critical approach and provide it to our interlocutors as a data set to be analyzed using computational tools of their choice. Then, we would debrief with the data scientists, asking them to describe the approach they had taken to the corpus and the conclusions they had drawn about the literature from it: did they accept or even understand some of the criticisms? How did they feel about

the qualitative methodologies underpinning these works? Transcripts of those debrief sessions would provide us with material for further analysis.

This plan immediately raised a set of conceptual problems: What corpus could represent the critical approach we identified with? Who would count as a suitable participant? Should we suggest tools in advance? These decisions raised typical STS concerns: What values and assumptions were embedded these decisions? How might those embedded values affect our intervention? While the tensions that inspired the project were clearly felt in our field-work encounters, how could we investigate them without inadvertently reifying or taking for granted the divide between data scientists and critics that we wanted to learn about? Rather than solving these problems through philosophical reflection, we decided to work in a prototyping mode—making a set of more or less arbitrary choices about corpuses, techniques, and participants, and then contacting people we knew from the field to see how they responded to the plan. Hence, we started with two potential corpuses, the Critical Algorithm Studies Reading list, assembled by Tarleton Gillespie and Seaver, and the back catalogue of *Big Data & Society* and decided to suggest some common forms of textual analysis like co-word (Callon et al., 1986; Danowski, 2009), topic modelling (Blei et al., 2003), and perhaps specific packages like like CorText which do both (<http://www.cortext.net/projects/cortext-manager/>).

Although we had ambitions to launch the exercise on popular data science forum Kaggle (<https://www.kaggle.com>), we settled on first emailing our colleagues in computer science departments. As responses to our emails came in, we realized that our experiment was already underway.

Email from Seaver

I’m working with a collaborator from Linköping University on an experimental project trying to facilitate communication between qualitative/critical researchers interested in data science/algorithms stuff and people working in those areas in a technical capacity[...]

[...] We’re in the process of figuring out our overall protocol, so nothing is set in stone yet. As far as sampling participants, we may stay with a convenience sample of a few people we know personally, or, if we can get a more coherent or geographically co-located group together, may go with that.

Whatever thoughts you have, I’d be interested in them! No urgency, though.

Thanks,
Nick

This message was answered promptly by an out-of-office auto-reply. Two months later, a response came in. The computer science professor apologized for the “epic delay” in responding and expressed interest in the study, which he framed as helping “...the more quant-heavy data sci people” to understand the critical literature and agreed this was important because of “gaps in terminology and epistemology” between the two sides. He said he was happy to assist or even participate, but felt that he was not “a terribly representative candidate” because he was already convinced of the need for better crossover between the fields.

The professor’s response speaks to some recurring observations about Computer Science and Data Science from our fieldwork. The delayed response is a reminder that computer scientists often have far more strenuous publishing schedules and are perhaps less available to engage in speculative and open-ended projects. More interesting, however, was his formatting of the problem as a matter of “terminology and epistemology”—that is, if we could find agreement on the meaning of words and underlying assumptions about knowledge, we could start to bridge the divide. This framing differed from certain currents of STS and anthropology, which assume that the “meaning of words” and “underlying assumptions about knowledge” are upshots of everyday practices, not the stable foundations on which they are built. So even articulating the problem raised disciplinary tensions.

Regarding the selection of participants, this respondent seemed to take on a statistical normative stance, perhaps in response to our use of the phrase “convenience sample,” wondering if he was properly “representative” of the group in question. This implied that we had defined these groups in advance, such that they could be adequately sampled, while we had hoped that the performed boundaries of different groups would emerge as an outcome of the experiment. Nonetheless, we realized that we might have necessarily been drawing on some caricature of a “quant-heavy” practitioner in order to solicit participants in the first place. The problem the comment raised was that because practitioners have various degrees of engagement with “critique,” it might prove impossible to find any participant who would be interested in participating and who was completely ignorant of critical work. It was possible that the setup of our experiment, by focusing on the literature, already presumes a divide between those who have been exposed and those who have not. Obviously, just as there are gradients in our field(s) between critical and complicit, philosophical and empirical, there are also gradients and overlaps within data science we need to consider. So even though we had not conceived of our intervention as a “controlled experiment,” we realized that some of our

underexamined assumptions about the field had smuggled their way into our initial prototype.

Another respondent, a graduate student in Computer Science observed that our proposed experiment was like “a snake eating its own tail,” which was not the only time potential participants remarked on the self-referential nature of the enterprise. He focused his attention not on epistemological underpinnings but on the more practical questions about the design of the “experiment.” He said that we should be careful about the tools and corpus we chose so that they did not “bias our result,” for example if we suggested co-word analysis, this might privilege a “network” or “community detection” view of the corpus. The observation that particular tools will steer the analysis and partially determine what we find, what becomes thinkable, might as well have been plucked from the critical algorithm studies literature itself. This response pushes against the caricature of data scientists as “unreflexive,” but suggests that the terms of this reflexivity may not match our own (Neff et al., 2017): he phrased his concerns in terms of “bias,” whereas we saw the formatting work of tools as an inevitable facet of research, which is itself empirically interesting. Indeed, we had hoped that our participants would be confronted by the limits of tools like co-word and topic modelling for analyzing dense academic prose and reflect on this. Still, the point stands that by encouraging participants to use a collection of “off-the-shelf,” well known tools, we might be preventing them from developing more creative solutions or even unexpected hybrids of digital and qualitative approaches. Our setup seemed to presume that participants would fail to analyze the texts (by our own interpretivist standards), and it perhaps did not grant them the possibility of succeeding (on their terms) to extract insights from and find patterns in the texts.

This respondent also started to draw more boundaries *within* “data science,” distinguishing text-specific techniques from other varieties: “We don’t do much of that in our lab.” We had focused on texts because this is the communications medium of choice for critical algorithm studies, but also because texts provide a classic object for “qualitative” interpretation. Therefore, while we might have imagined our ideal participant as extremely number-oriented (or “more quant-heavy,” as the first respondent put it), these responses suggested that we should in fact looking for people who appeared disciplinarily “closer” to us, possibly in a field like digital humanities, who specialized in deriving meaning from large amounts of recalcitrant textual data. These and other responses to our initial idea brought into relief some of the contours of the fields we were interested in—including normative positions on what counts as an “experiment,” “bias” or “success”—and they

also probed our own assumptions about who would be an “ideal” participant, given that the divide was far more fluid and variegated than we had anticipated. Even though the experiment had not yet begun, the gap we had identified was already taking on a texture.

The halting nature of these communications, however, also suggested another interpretation: we could read these responses as tactical efforts at deferral, avoiding commitment to a project that seemed vague and/or confusing, but also avoiding straight rejection. In other words, the respondents’ redefinition of “the problem” and suggestions of the salient types of people who might be involved in exploring it were not only drawing on ready-to-hand common-sense tropes from their own social worlds like statistical representation or experimental bias or making strong statements about the nature of disciplinary identity—we could also understand them as attempting to avoid direct involvement while passing the experiment on to others. In any case, while we received several polite and lengthy responses to our proposal, this version of the experiment never really got off the ground.

Having little success in staging the experiment as we had originally intended, we decided to change tack. Perhaps, our aim should not be to provoke some hardened, objectivist data scientists (if they actually exist), but instead to examine one situation from which such a figure might emerge: in the classroom. New data science programs are popping up at universities around the world, and in some cases, people we might have considered “critics” have become involved in designing their curricula—sociologists, historians, ethicists, and other STS scholars have worked to overcome the gap we had identified by intervening here, in the educational process. Since data science as a practical and educational field is still in formation, asking students to participate in our experiment would not only test the qualities of this apparent dichotomy but also intervene in a site of its potential production.

We refigured our plan as a classroom exercise⁴ and sent it to social scientist colleagues involved in creating curricula for data science programs. These scholars were often tasked with representing “ethics” in their respective programs—a typical location for humanistic and social scientific enterprises within institutional scientific settings. A sociology professor responded enthusiastically to our proposal and CCed colleagues in Human Computer Interaction and Computer Science. While they were developing a data science masters, in the meantime they suggested that students from a Python training workshop could work, even though they were “not your typical data scientists.” Again the invocation of “typical” raised questions about who we thought we were looking for—what would a future data scientists look like?—and forced us to

acknowledge the tensions between representing aspects the divide “faithfully” and simultaneously trying to change it.

Unsurprisingly, these respondents were much more receptive to the second prototype of the experiment, perhaps because we identified the right “frictions” (Zuiderent-Jerak and Bruun Jensen, 2007) and found a way to make the experiment “useful” for those involved. In this round of contacts, we had also begun with closer disciplinary “kin” in the social sciences. Although we continued to encounter deferrals, these deferrals were more productive than the ones described above, moving us through social networks from close contacts to more distant associations. However, one CS professor, who also responded favorably to the idea, commented that it was “incredibly meta!” Similarly when one of us pitched the experiment verbally to a professor in Human Computer Interaction, the inevitable reply was “you social scientists love mind games.” These references to circularity raised further questions about the experiment and why it seemed to read as indulgently self-referential for different disciplines, which we will reflect on below.

Encounter two: “It’s not Woodstock”

In the wake of Cathy O’Neil’s editorial, we brought these initial observations about the experiment to the aforementioned Algorithms Network meeting, looking for feedback from fellow STS scholars of algorithmic systems. Here, to our surprise, we had another encounter that drew into question our formulation of the problem. Although we had taken ourselves to be representatives of the STS approach, encountering a set of expected resistances and reframings in the world of data science, we found that even on our “home” turf, our planned intervention elicited unanticipated boundary-marking efforts among STS scholars. That is to say, we were learning not only about how data science practitioners understood and enacted the gap between their own work and that of their interpretive critics but also how critics themselves, variously committed to projects of identifying, reconciling, and communicating differences, enacted that gap themselves. What had begun as an ordinary effort at collecting feedback from peers quickly turned into a spontaneous moment of fieldwork—an unanticipated “para-site” (Marcus, 2000) for investigating the questions that had motivated our project from the start. In these responses, we found a tension between, on one hand, arguments that we had exaggerated the divide between data scientists and their critics and, on the other hand, those that suggested the divide was even deeper than we had described.

In the first instance, many of our fellow social scientists seemed to suggest that we were overstating the

divide between practitioners and critics. Although we understood ourselves to be presenting an empirical tension, encountered in our ethnographic work, some took us to be reinforcing an “us vs. them” mentality, suggesting that we had brought this frame to our work, rather than finding it there. In our casual use of terrain metaphors to describe the interfacing of disciplines (borders, territory, and so on), we were taken to be using the language of war, as though the two sides were lobbing mortars over trenches.

However, other participants who had experience of collaborating with data scientists confirmed the existence of a divide. While acknowledging that the past few years have seen several fruitful collaborations, one remarked, “it’s not Woodstock”—not an idyll of positive feeling and mutual acceptance. Interestingly, this same participant pointed to a longer disciplinary history, in which techniques like natural language processing could be connected more directly to humanistic disciplines. “It wasn’t always like this [i.e. disciplinarily isolated], we have been segregated,” he said, noting that newer researchers in these fields, more interested in machine learning, tended to shut out humanistic theorists and social scientists. Another participant hearkened back to the earlier tensions between STS scholars and their scientific counterparts, remarking that “It’s not a war like the science wars, but what is it now?”; she analogized the present moment, and our problem, to a post-conflict situation, where people from formerly feuding groups had to find a way to “make the peace.” Others pointed to what they suggested were fundamental divisions within data science (and statistics more broadly), between the study of language and numbers or frequentist and Bayesian approaches.

These efforts to locate our apparent “divide” in historical trajectories seemed to simultaneously reify it, as an unavoidable fact of the present situation, and to denaturalize it, as an historical contingency that could be undone or remade differently. Perhaps, the division was less like a contested borderland and more like the division between two moieties in a kinship system, who could trace their lineages back to common ancestors but nonetheless found themselves structurally at odds in the present. (Seaver has advanced such an argument, suggesting that scholars attempt to study “the kinship of methods” 2015).

What quickly became clear was that, even within STS, responses to our intervention were diverse and contradictory. Just as we had observed that there are many conflicting versions and gradations of data science, we were now encountering different camps within “our side,” with various empirical and theoretical commitments. This simple observation has significant consequences, which our planned intervention and the responses to it tended to elide: varieties of disciplinary

ways of knowing can significantly impact the way we conceptualize the divide, whether we identify it as a divide in the first place, and what routes we might take to overcome it.

It seemed to us that some of the responses could be explained by a tendency among STS researchers to assume that this divide, like all dichotomies, needed to be thrown on the “bonfire of the dualisms” (Law and Hassard, 1999). But, in the process of recognizing that divisions are constructed, contingent and not natural, this response downplays actually existing tensions and boundary work, which participants who had worked with data scientists, knew all too well. These tensions cannot simply be wished away. Hybrids between critical algorithm studies and data science are not what remains when boundaries are removed, they need to be built up, practice by situated practice, on sometimes shaky ground.

Just as our first encounter made plain the agreed-upon assumptions and baggage underpinning how data scientists conceptualized the divide (and the experiment) this second encounter made clear, how we also needed to unpack our own disciplinary assumptions “within” STS in order to understand how they might shape our understanding of the divide and our ability (or desire) to talk across it. What this encounter made clear was that different traditions within STS and critical algorithm studies which are committed to: showing the historical contingency of knowledge, breaking dichotomies, and recasting social phenomena as situated practice shape the way we conceptualize the divide in different ways. In other words, to perhaps state the obvious, even the position that normativities are the upshot of practice, is itself a normative position which has implications for what realities we enact.

Reflexivities

This encounter left us in somewhat of a quandary. We had confirmed that the divide was real but much more complicated than we had originally imagined. And yet, our attempts to communicate this complex tangle of practices seemed to consistently confuse our data science interlocutors (as well as some from our own discipline). How could we address this difficulty in communication? Michael Lynch (2000) writes about the many versions of reflexivity—from those which are intended to make activities more objective to those which draw attention to the constructed or situated nature of practice. For ethnomethodologists, as Lynch argues, reflexivity is an ordinary background feature of *all* interactions—the means by which people make themselves account-able to others. What is significant then is: in what situations and in what ways, it is permissible to refer to the accounts themselves? We

noted earlier that the caricature of data scientists as unreflexive is not accurate, but the qualities of their reflexivity differ from our own. Data scientists, it might be claimed, practice the form of reflexivity Lynch calls “Methodological Self-Criticism” in which a certain version of objectivity is achieved by systematically laying out one’s biases on the table and adjusting for them. However, it could be argued that anthropologists are more inclined to practice the subtly different form of reflexivity Lynch names “Methodological Self-Consciousness,” in which the researchers’ positionality vis-a-vis the groups phenomena being studied is explored.⁵

For example, many of the ethnographers in the room felt that we were being overly critical when we described our data science participants, or that they were “drawn too simply.” This style of reflexivity is concerned with the obligation to speak on behalf of the people we are studying, to do justice to our informants, often because the objects/subjects of ethnography are seen as not having a loud enough voice. The irony in this particular situation is that data scientists generally have a much louder voice than us anthropologists. A related disciplinary move which might also count as a mode of reflexivity is the generalized symmetry principle (Callon, 1984) which insists that researchers explain phenomena across established dichotomies: true/false, human/non-human, nature/culture, using the same conceptual equipment. Many of the participants argued that we were not being sufficiently symmetrical about the study. One participant proposed that in addition to the critical data studies papers, we should also analyze data science papers using a similar approach or that we should include our own emails as part of the analysis. Framing research in symmetrical terms is one key way that STS researchers make themselves accountable to each other.

Symmetry is a theoretically appealing principle and can be a productive tactic for careful observation, but it may be less useful for thinking about and planning interventions. Symmetry presumes that the researcher is in some sense in control of the research, that they can position themselves as a fulcrum to balance certain dualisms. However, we saw that the field we are trying to intervene in itself is asymmetrical: as some have experienced in Data Sprints and Hackathons—some actors have more resources and this means that we cannot necessarily intervene in the way that we choose. This manifests itself as a tension between being fair in our representations of data scientists, while also sticking up for the validity of our own somewhat marginalized position.

In any case, these forms of social science reflexivity may be more or less productive analytically, but the question which Lynch’s account raises is: How

do these forms of reflexivity come across to our data science counterparts? Are our ways of becoming “account-able” at all legible to our interlocutors? As we saw in the previous section, there were multiple references to “snakes eating tails,” “meta” exercises, and “mind-games” so it seems that our research subjects read our intervention as examples of the style of reflexivity Lynch calls “Breaking the Frame” that is, the artistic practice of drawing attention to the constructedness of a medium (like “breaking the fourth wall” in a film)—or maybe “Reflections ad infinitum”—the *mise en abyme* technique of endless self-reference—something we are doing now by reflecting on the ways we reflect on things. Such styles of reflexivity are viewed as a sign of intellectual rigor or “cleverness” in some academic circles, while being viewed as pretentious or navel gazing by others.

While the paper you are reading is not innocent of navel gazing, such an exercise can be worthwhile so long as it results in learning and changes in practice (Woolgar, 1988). The serious point to be made here is that part of the divide between data scientists and their qualitative critics has to do with subtle differences between how the two camps (and divisions within those two camps) become accountable to each other. Thus ironically, as one of our data scientist respondents said, perhaps we do need to address the divide through language. The problem, as we identified, is that terms like “reflexive,” “experiment,” and “representative” mean different things in different communities and rather than papering over these differences in “shared” language we need to explore the differences and topicalize these for our informants. This was the idea behind our revised classroom exercise, which invites a generative confrontation between different ways of knowing but does so in a familiar form, embedded in existing networks and routines. We have made an example protocol available online (<http://www.nickseaver.net/s/algorithm-studies-exercise.pdf>) so that readers of this journal who find themselves in this crossroads between disciplines can take up the experiment and modify it as they see fit.

Conclusion

There have many valuable attempts at collaboration between social scientists and programmers/data scientists in recent years, but as we have argued, many of these partnerships accept normal disciplinary divisions of labor: social scientists observe, data scientists make; social scientists do ethics, data scientists do science; social scientists do the incalculable, data scientists do the calculable. While some of these initiatives are starting to take hold, fundamental misunderstandings

between these disciplines remain, as the fall-out over Cathy O’Neil’s editorial makes clear.

In this paper, we have argued that shaking up these disciplinary divisions might require a more experimental, interventionist approach in order to better draw out the normative commitments which sustain these tensions. We proposed an experiment, designed to probe the divide between data scientists who make algorithms and qualitative social scientists who study them by encouraging them data scientists to reflect on the content of these criticisms and in turn the approaches they use to make sense of them. We recounted our sometimes fumbling, grasping attempts to learn about how data scientists conceptualize the divide, if at all, but in the process ended up learning as much about our own (mis)conceptions of the field. This is because experiments (of the open ended *or* controlled variety) force us to make choices and commit to assumptions in ways which are less common in so called “qualitative” disciplines (but which are an everyday feature of data science work). Although our intervention has barely started, it has already highlighted several points of tension and potential misunderstanding.

While we cannot make strong claims from these few encounters alone, they resonated with previous observations from our fieldwork: that often the critical literature on data scientists paints them in simplistic ways; that they are far more critical of their own tools and acknowledge doubts about the limits of computational analyses on difficult objects like texts. We also saw that there are many hybrids of data science and the critical position—there is no “no man’s land” but rather complex networks of practitioners who nonetheless invoke distinctions and boundaries in everyday talk. For this reason, we found the experiment to be more successful, though perhaps less provocative, when we embedded it within existing networks and attachments.

We also learned about some of the tacit assumptions and normativities in the data science field about the nature of experiments, even playful ones like ours. But through feedback from our social science peers, we also learned about how our own inherited (and sometimes conflicting) disciplinary assumptions format our ability to understand and navigate the divide. For example, our own commitments to locating conceptual work in situated practice and to treating divides symmetrically may hamper our ability to intervene in ways our informants understand. In particular, we highlighted that we need to understand better how our multiple versions of reflexivity, how being “accountable” to our own disciplines, might not always be legible to our counterparts. In other words, any attempt to rethink disciplinary divides means that our own normativities need to be put at risk and sometimes bracketed.

To return to Cathy O’Neil’s article, we should acknowledge that at least some of the lack of recognition is the fault of critical algorithm studies scholars. We need to find better ways of becoming interesting or useful to our data science counterparts without, on one hand, adopting wholesale their terminology of “ethics” and “bias” or, on the other hand, leaving unexamined our own qualitative or critical frame. What is needed are more studies of the divide that do not paint it as a fiction but also do not take it as inevitable.

Acknowledgements

The authors would like to thank our data science interlocutors, the Algorithms Network for prompting this experiment (and participating in it) and the Values group at Linköping University for helpful feedback on an earlier version of the text. In particular, we would like to thank Ivanche Dimitrievski for pointing us to Lynch’s work on reflexivity and to the anonymous reviewers for some helpful literature suggestions.

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. Indeed, this framing is so familiar that it has a name: Malte Ziewitz, in the introduction to a 2016 special issue of *Science, Technology, and Human Values* dedicated to the topic, called it “the algorithmic drama.”
2. Many of these research agendas call for forms of artistic practice, designing software tools, or more institutional engagement exercises like toolkits and playbooks: in other words anything but the humble academic journal article or theory-heavy monograph. But, it is also important to remember that textual accounts are also, already interventions (Vikkelsø, 2007), in the sense that they enact particular ontologies (Mol, 2002) which may interfere with accepted normativities and situated practices. Yet, regardless of whether we are talking about academic texts, official reports, flowcharts and diagrams or material, designed artifacts, we need to consider possible frictions—which practices can we graft our interventions on to (Zuiderent-Jerak, 2007).
3. Interventions may in this sense work as “breaching experiments” (Garfinkel, 1963) which reveal the “background expectancies” which underwrite social order as a local accomplishment. However, as Garfinkel (2011) notes breaching experiments are more *demonstrations* of the existence of such expectancies and orderings, rather than attempts to either influence or systematically learn about particular social arrangements.

4. An example handout for the exercise can be found here (<http://www.nickseaver.net/s/algorithm-studies-exercise.pdf>).
5. Lynch lists six types of reflexivity with multiple sub-branches each.

ORCID iD

David Moats  <http://orcid.org/0000-0001-9622-9915>

Nick Seaver  <http://orcid.org/0000-0002-3913-1134>

References

- Adams V (2016) *Metrics: What Counts in Global Health*. Durham, NC: Duke University Press.
- Blei DM, Ng AY and Jordan MI (2003) Latent dirichlet allocation. *Journal of Machine Learning research* 3(Jan): 993–1022.
- Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.
- Burrell J (2016) How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1): 1–12.
- Callon M (1984) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review* 32(1): 196–233.
- Callon M, Law J and Rip A (eds) (1986) *Mapping the Dynamics of Science and Technology*. London: Springer.
- Callon M, Millo Y and Muniesa F (2007) *Market Devices*. Hoboken, NJ: Wiley-Blackwell.
- Danowski JA (2009) Inferences from word networks in messages. In: Krippendorff KH and Bock MA (eds) *The Content Analysis Reader*. London: Sage, pp. 421–429.
- Downey GL and Zuiderent-Jerak T (2016) Making and doing: Engagement and reflexive learning in STS. In: Felt U, Fouché R, Miller CA and Smith-Doerr L (eds) *The Handbook of Science and Technology Studies*. Cambridge, MA: MIT Press, pp. 223–252.
- Garfinkel H (1963) *Studies in Ethnomethodology*. Oxford, UK: Blackwell Publishers Inc.
- Garfinkel H (2011) A conception of and experiments with “trust” as a condition of concerted stable actions. In: O’Brien J (ed.) *The Production of Reality: Essays on Social Interaction*, 5th edition. Thousand Oaks, CA: Pine Forge, pp.379–391.
- Gieryn TF (1999) *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University of Chicago Press.
- Gitelman L (2013) *Raw Data Is an Oxymoron*. Cambridge, MA: MIT Press.
- Hall S, Hobson D, Lowe A, et al. (1980) Encoding/decoding. In: Hall S (ed.) *Culture, Media, Language: Working Papers in Cultural Studies, 1972–79*. New edition. London: Routledge, pp.117–127.
- Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 1–7.
- Irani L (2015) Hackathons and the making of entrepreneurial citizenship. *Science Technology and Human Values* 40(5): 799–824.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Law J and Hassard J (1999) *Actor Network Theory and After*. Oxford, UK: Blackwell Publishing.
- Lazer D, Pentland AS, Adamic L, et al. (2009) Life in the network: the coming age of computational social science. *Science* 323(5915): 721.
- Lipton C and Steinhardt J (2018) Troubling trends in machine learning scholarship. Paper presented at ICML 2018: The Debates, Stockholm 27 July 2018.
- Lynch M (2000) Against reflexivity as an academic virtue and source of privileged knowledge. *Theory, Culture & Society* 17(3): 26–54.
- Marcus GE (2000) *Para-sites: A Casebook Against Cynical Reason*. Chicago: University of Chicago Press.
- Marres N and Weltevrede E (2013) Scraping the social? Issues in real-time social research. *Journal of Cultural Economy* 6(3): 313–335.
- Mol A (2002) *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.
- Munk AK, Tommaso V and Meunier A (2019) Data Sprints: A Collaborative Format in Digital Controversy Mapping. In: *Digital Sts Handbook*. Princeton University Press.
- Neff G, Tanweer A, Fiore-Gartland B, et al. (2017) Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data* 5(2): 85–97.
- Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.
- O’Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.
- O’Neil C (2017) The ivory tower can’t keep ignoring tech. *New York Times*, 14 November. Available at: www.nytimes.com/2017/11/14/opinion/academia-tech-algorithms.html (accessed 5 June 2018).
- Osborne T and Rose N (1999) Do the social sciences create phenomena? The example of public opinion research. *The British Journal of Sociology* 50(3): 367–396.
- Pasquale F (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- Rabinow P, Marcus GE, Faubion JD, et al. (2008) *Designs for an Anthropology of the Contemporary*. Durham, NC: Duke University Press.
- Ruppert E, Harvey P, Lury C, et al. (2015) Socialising big data: From concept to practice. *CRESC Working Paper Series* (138). Available at: <http://research.gold.ac.uk/11614/> (accessed 24 November 2016).
- Schutt R and O’Neil C (2013) *Doing Data Science: Straight Talk from the Frontline*. Sebastopol, CA: O’Reilly Media, Inc.
- Seaver N (2015) Bastard Algebra. In: Maurer B and Boellstorff T (eds) *Data, Now Bigger and Better!* Chicago: Prickly Paradigm Press, pp. 27–45.
- Stengers I (2000) *The Invention of Modern Science*. Minneapolis: University of Minnesota Press.
- Uprichard E (2013) Focus: Big data, little questions? *Discover Society*. Available at: www.discoversociety.org/2013/10/

- 01/focus-big-data-little-questions/ (accessed 30 September 2014).
- Vikkelsø S (2007) Description as Intervention: Engagement and Resistance in Actor-Network Analyses. *Science as Culture* 16(3): 297–309.
- Wallach H (2015) Conflusion – Computational Social Science: Toward a Collaborative Future. In: Michael Alvarez R (ed.) *Data science for politics, policy, and government*. Cambridge, UK: Cambridge University Press, pp. 307–316. Available at: <http://dirichlet.net/pdf/wallach15computational.pdf> (accessed 3 December 2018).
- Wilkie A, Savransky M and Rosengarten LSM (2017) *Speculative Research: The Lure of Possible Futures*. UK: Taylor & Francis.
- Woolgar (1988) *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*. London: Sage.
- Ziewitz M (2016) Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values* 41(1): 3–16.
- Zuiderent-Jerak T (2015) *Situated Intervention: Sociological Experiments in Health Care*. Cambridge, MA: MIT Press.
- Zuiderent-Jerak T and Bruun Jensen C (2007) Editorial introduction: Unpacking ‘intervention’ in science and technology studies. *Science as Culture* 16(3): 227–235.