

Stephen Paff

Spring 2018

The Anthropology of Machine Learning

Introduction

Machine learning algorithms are a recent development in the last few years as methods in which machines learn through iterative algorithmic processes based on data or other inputs to form knowledge. They are a prime set of entities within the broader field of data science, an interesting, decentralized discipline in of itself. Recent developments within and interest in machine learning and data science have received significant support from increasing interest in supposed “big data” and desire to develop large-scale, software data analytical infrastructure within the business world and academia (Kitchin 2014:xvi-2, Kelkar 2014, Intron 2016:27, Carden 2016:102). All of this complicates both a basic understanding of what machine learning, data science, and big data are, and they also feed significantly into how anthropologists and other social researchers from related fields like sociology, science and technology studies, human-computer interaction, design, etc. have approached these sociotechnical phenomena.

As someone with both a background in both anthropology and mathematics, with formal trained in some of bodies of knowledge typically considered canon for machine learning and data science, and also trained in many of the anthropological theories invoked to analyze these phenomena, I have, at times, found myself uniquely vexed by the conversation. In many ways,

this paper is my attempt to sort this out. As someone who is fluent in both sides, I have seen both parties talk past each other for a very long time, and this paper in significant ways is a call to actually stop and listen to what the other is saying. Writing this paper primarily for anthropologists, I am seeking to articulate what I see as the shortcomings of this “side” and what I see it as needing to do to understand the “other.” If I wrote this paper primarily for a data science, mathematics, or computer science audience (a future endeavor which may come to fruition one day), I would approach it from a different angle.

The hype surrounding these concepts have led to exaggerated and, at times even, fictionalized depictions. With that in mind, for the rest of this introduction, I hope to define *machine learning*, *data science*, and *big data* as well as what the *anthropology of machine learning* could mean, and why I decided to write a paper on it. First, I will show the interrelationship between three concepts and define them formally. Embedded within each definition will include definitions of related words: for *machine learning*, defining *knowledge production*; for *data science*, defining a *data scientist*; and for *big data*, defining *data*. Then, I will develop the *anthropology of machine learning* and its importance. All of this will, of course, culminate in the introduction of a thesis for the essay as a whole.

The paragraphs above described machine learning, data science, and big data as connected sociotechnical phenomena. They are not the same,¹ yet they do connect heavily. The following statement best summarizes their interrelationship:

A: *Data scientists study (big) data through machine learning (primarily).*

¹ This is despite the frequent false equivocations I have heard among quotidian descriptions, including, at times, by many anthropologists (generally not anthropologists who are experts in the subfield, at least, who tend to know better).

A indicates the who or practitioners – data scientists – the what or subject matter – data, including big data – and the how or methodology – machine learning. The following is a parallel summative oversimplification for anthropology:

B: *Anthropologists analyze culture through ethnography (primarily).*

Each are an oversimplification, represented by the inclusion of *primarily* at the end but is useful as a rough model.² Technically, not all computational techniques that data scientists use classify as machine learning and not all problems that data scientists and/or machine learning seek to solve classify under big data.³ As a matter of fact, data scientists, using both machine learning and other strategies, do not only strictly work on big data problems, but often on problems of all sizes, which may or may not connect or have anything directly to do with the types of datasets, work, and inquiries typically labelled as “big data.” It holds weight as a rough way to express the interrelationship between the four big terms (data science, data scientists, machine learning, and big data) used to describe the phenomena computational data analysis: the field of data sciences represents a discipline in which data scientists analyze big data through machine learning.

Since demonstrating the interrelationship between these three, sociotechnical phenomena, I will now define each one: machine learning, data science/scientists, and big data. Each has multiple different definitions and perspectives on what they mean, which complicates easy description of their meanings, yet all of them have a basic, common working understanding.

² The following two statements would be more accurate: “Data scientists (primarily) analyze (big) data (primarily) through machine learning (primarily)” and “Anthropologists (primarily) analyze culture (primarily) through ethnography (primarily)”. I simply put *primarily* in the sentence only once instead of placing it at every spot where it is technically necessary for the sake of visual clarity and readability.

³ In parallel, not all anthropologists use ethnography or analyze *culture*, and anthropologists are not the only people using ethnography.

Machine learning generally refers to a set of algorithms, which *learn* from a set of data. I would formally define an algorithm as *machine learning* if it “learns” by developing its own way to analyze data (in the same or new context) by adapting over iterations. Data scientists have increasingly developed a “canon” of common types of algorithmic processes for data analytics, which the phrase *machine learning* typically connotes, including decision tree modeling, neural networks, deep learning, logistic regression, collaborative filtering, support vector machines, cluster analysis, etc. Now, as one can tell by the phrase, what it means for a computer/machine to learn is extremely important for understanding what a machine learning systems is. For computers/machines, “learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time” (Kononenko 2007:37, quoting Herbert Simon). In short, machine learning iterate over data and in each iteration refine how it analyzes the data, which are typically used to develop some model from the data.

Within my understanding of computers learning is an implied definition of knowledge, given that in many cases, these computer systems develop knowledge through their learning. What does it mean to say something produces knowledge? Knowledge has many definitions, both in everyday contexts and by scholars, yet most definitions of knowledge production center around one or both of these two concepts: developing/gathering facts and information and gaining skills, awareness, or familiarity. The former emphasizes concepts gleaned; the latter embodied dispositions, affordances, or intuitions, which manifest as actions. For example, consider when Google searching the word “knowledge” the following two definitions resulted:⁴

⁴ I searched “knowledge” on Google at 12:20 pm (U.S. Central time) on April 25th, 2018.

1. “Facts, information, and skill acquired by a person through experience or education; the theoretical or practical understandings of a subject”
2. “Awareness or familiarity gained by experience of a fact or situation”

The second definition explicitly emphasizes the type of awareness or familiarity, which emphasizes embodied states like awareness or familiarity, which presumably manifest in (correct or successful, by whatever understanding of these used) action.⁵ The first emphasizes concepts of facts and information (although also including skills, which is arguably closer to an embodied state).⁶

For machine learning, both definitions of knowledge apply. The algorithms develop facts or information, such as vectors, weights, Boolean structures, etc., as it iteratively analyzes the data. It then uses these facts and information to conduct some actions or set of actions on the data and/or on other data, potentially including predicting values in the data, classifying values, recommending actions for other agents, and adjusting its hyperparameters (a.k.a. adjusting how it adjusts to more data in the future), explicitly using these facts/information to adapt or change how it analyzes the data. Through its analysis of the data, machine learning algorithms shift the process they use to analyze future data (whether that be by developing a Boolean decision structure, adjusting the values in an decision-making equation, adding or adjusting layers in its

⁵ Within both definitions is an implied positive trajectory of acquisition, the skills, awareness, and/or familiarity, a.k.a. knowledge produced improves the learner’s ability in some way. In many colloquial discussions of producing knowledge, this is implied, but I do not consider it strictly necessary: knowledge is produced when it leads to embodied dispositional changes or adaptions. Although often pertinent in assessing the quality of the knowledge formed, whether those adaptions prove ultimately beneficial or harmful (or somewhere in between) is not strictly a criteria for whether knowledge formation has occurred.

⁶This definition formally articulates persons (an extremely complicated concept theoretically in of itself) as the subject of knowledge production through its phrase “by a person.” In Section 3, I articulate an understanding of agency and of agent, and I would define any agent as the potential subject for knowledge production, which would include non-human, specifically computerized agents. In this essay, I am explicitly invoking the term *agent* to define the various subjects involved in knowledge production, and thus what constitutes a person is not necessary for this paper. Suffice to say succinctly, however, I consider the words *agent* (as I define it in Section 3) and *person* to mean the same thing, but I prefer the term agent because it has less associations with exclusively human persons.

decision-making, etc.), and thus machine learning constitutes a form of skill or familiarity building invoked within conventional understandings of knowledge production.

Data science is an interdisciplinary or intersectional discipline that seeks to analyze and extract data, typically through the development of computational algorithms. A friend of mine and fellow data scientists provided the most basic yet most compelling definition of data science I have heard: “Data science is applied, computational statistics.” Situating itself in terms of the analyzing data means that data science technically falls within or at least close to statistics (which at the most basic level is the mathematical analysis of data),⁷ yet data science puts the emphasis on computational statistics: developing/programming algorithms to analyze data. At the same time, unlike academic or “pure” computational statistics (or academic statistics in general), data science has evolved in the context of analyzing specific data and solving “real world” problems. This friend quipped that data science is the term for computational statistics used outside of non-academic contexts, particularly the business world, which illustrates pertinent aspects of undergirding relationships. First, its most immediate driving force came from outside of the academic world, yet it is becoming incorporated into academia. Second, computational strategies is foundational to the practice, with significant influence from computer scientists in the field.⁸ Finally, data science situates itself as a science (which is in the name itself), yet it also defines itself differently than “conventional” scientific practices. As the sociologist/anthropologist Shreeharsh Kelkar (2014) demonstrates that data science both reflects

⁷ My own experience has taught me that when anthropologists and many social researchers think of statistics, they conceive of very specific sets of practices (like certain hypothesis testing methods), which data science can involve but does not have to.

⁸ This is unlike statistics within mathematics, which now definitely has a significant computational aspect, yet which also experienced formative development before the creation of personal computers.

and catalyzes a shift within the hard sciences towards computational analytical work, requiring “a certain kind of scientist who is skilled at both statistics and software-building.”

Both as a unique mixture of several facets, a fairly decentralized, loosely connected set of theories, methodologies, techniques, and approaches, a newness in which the discipline is still in the process of significant formation, it is a pretty hard discipline to pin down formally in a definition, yet practices and techniques within have been standardizing/solidifying. Nick Seaver, an anthropologist in this area, defines as “data scientists work as professional bastard-makers, combining data sets, algorithms, and epistemologies in unauthorized ways to produce illicit offspring” (2015:43). Referring to the discipline as a bastard (which he did not intend as an insult) reflects how the discipline is a unique hodgepodge of different “people, epistemologies, and methods” (including mathematics, computer science, engineering, economics, sociology, etc.) brought together to solve a specific set of problems: analyze datasets (2015:43). This parallels anthropology, which (re)developed significantly in the 1920s and 1930s as a bastard set of practices for analyzing cultures loosely called ethnography (2015:44). Just like anthropology has undergone solidification and formalization into a single, encapsulated discipline, data science may undergo something similar in the next several decades as well (2015:393-394, 398).

Even though it is clearly related to *data science*, understanding what a data scientist is an extremely important word as well, at the very least because data scientists themselves spend quite a bit of time defining themselves. In this paper, I will user the oversimplified definition of *data scientist* as someone who practices data scientist typically occupationally, in spite of the complexities of trying to directly relate what it means to “practice data science.” Trying to define how data scientists understand what it means to be a data scientist is a fascinating and necessary ethnographic project in of itself but beyond the scope of this paper. As a parallel, Keri Brondo

(2012) conducted a fascinating study into how anthropologists defined what it meant to be an anthropologist, which illuminated complexities on what it meant to do anthropological “work:” among those with graduate degrees in anthropology, those within and outside of academia these definitions possessed very different understandings of what anthropological work was and what it meant to do those things, with significant implications for how anthropology was defined as a discipline. Being conceived of as non-academic primarily and then academic secondarily, data science may reverse anthropology as a discipline (or at least contemporary conceptions of anthropology, c.f. Seaver 2015a), yet similar research could yield some parallel types of complexities.

Furthermore, complex performativity imbues this basic definition of a data scientist, centered principally on what it means to *do* data science. Similar to Jasper’s ethnographic analysis of Goths in Amsterdam who believed that true Goths would never call themselves Goths (2004:2), data scientists themselves often indicate that those who profess themselves are not real data scientists. This all becomes complicated given the use of data scientist as a job title, with a varying but somewhat defined set of potential practices and prestige associated with it. Understanding how definitions of data scientists and data science influence each other and how the performativity aspect of both as a marker of embodying a certain set of values and dispositions and as a role/title significantly influences what it means to “do data science,” requiring much more elaborate ethnographic research to articulate, yet the basic, simplistic structure of a data scientist is someone who does data science is still valid within these more complicated frameworks.

Big data is the last major term in the set of sociotechnical phenomena this paper is analyzing. Defining data that is supposedly big requires understanding what data is. I understand

data as information (characteristics, features, etc.) collected and analyzed (c.f. Kitchin 2014:2-4, 9-10). “Data do not exist independently of the ideas, instruments, practices, contexts, and knowledges used to generate, process, and analyse them (2014:2). In the context of data science, data typically represent mathematical objects (numbers, vectors, variables, etc.) organized into sets (c.f. Mackenzie 2012ab) stored computationally, constituting inputs and outputs of and often undergoing transformation within computational analysis. In this specific word, data constitute a the brick and mortar building blocks of analysis, but specifically in the context of data science, have a set of guidelines for how to produce, structure, relate, and distribute (Kitchin 2014:1).

There are a lot of different definitions of “big data,” most of which tries to delineate what makes a particular instance of “big.” I am less interested in definitions what makes such data big (which typically relate to constantly shifting, technical understandings of what the frontiers of size is for computational data and data storage anyways). I see *big data* as a tagline or buzzword to refer to the set of developing techniques or strategies which seek to collect, organize, analyze, and utilize (and in some cases transfer) large amount of data computationally. As such, the literal size of the data is less important to understanding big data and determining what constitutes a supposed example of *big data* than to understand the methods of collection, analysis, and use employed.

One end of the spectrum, these techniques represent a way to develop, organize, manage, and understand large datasets, facilitating the use of these datasets to produce useful and unique insights (Kitchin 2014:1-2, 12, Domingos 2015:xiv-xv). On the other end, these techniques represent a new way to control and regulate human individuals (Cheney-Lippold 2017:99). Although there is a wide variety of middle ground within this dichotomy and alternatives beyond it, this debate fuels an intense interest on the concept of big data within the general public. This

debate has significantly altered attempts to define and articulate what these techniques are. For example, within the hard sciences, Kelkar argues that “the anxiety around big data is less about a crisis in epistemology (the role of theory, ideas about statistical significance) and more about a crisis in professional identities of practitioners who work with data” (2014), a sentiment that I will show is common among anthropologists and other social researchers in the first section of this paper.

After defining in detail what each term means, I will explain what I mean by the *anthropology of machine learning* and why I think that is something we as anthropologists need to develop. The concept may seem counterintuitive, particularly since anthropologists generally analyze cultures, and machine learning is not a culture in any sense of the term. Rather machine learning is a conglomeration of algorithmic methods used to solve problems, often embodied in specific code, which has a particular function in a set of overall code. These algorithms and overall process of organizing them within the concept of *machine learning* clearly connects with sociocultural entities, and a social analysis of them would be beneficial. Additionally, these sets of algorithms, called machine learning algorithms, are social entities in of themselves, with increasing sociocultural influence. As they seem to become more popular and thus created more and more to solve problems, they are becoming an increasingly important feature in our lives. Whether this trend continues (which is hard to tell), they currently exist. Thus, as anthropologists, we should seek to understand how they function as a part of cultures/societies (or *sociocultures*). The anthropology of machine learning is the word I am using for this latter understanding, also an explicit move away from the anthropology of big data and other former ways anthropologists have attempted to conceptualize this wider set of social phenomena.

In this essay, I will review the anthropological and other related social research of machine learning and related topics (such as data science and big data) to demonstrate that we, anthropologists and other social researchers, need to do the following:

1. To understand what machine learning, data science, and big data look like on the ground through ethnography, specifically by resituating ethnographic research towards machine learning and data scientists as opposed to big data.
2. To determine how to bridge the gap between large scale machine learning and other data science development and ethnographic inquiries.
3. To shift our understandings of who or what participates in a culture and society as increasingly intricate computerized systems develop their own knowledge.

Section 1: Ethnography of Machine Learning and Data Science

In this section, I will argue that we social researchers need to understand what machine learning, data science and data scientists, and big data look and feel like on the ground, and that a move away from big data towards ethnographically an understanding of machine learning and data science would help drive this. Many anthropologists and other scholars have laid the groundwork for this type of work, but the current overall research and literature on the topic reflects a need for more on-the-ground, everyday understanding of such practices. In this section, I will review the literature to both demonstrate the need and to demonstrate several researchers whose methodological and theoretical developments could provide a launching point for this type of ethnographic focus.

Searches in the anthropological literature for the three big terms of this topic – data science, machine learning, and big data – reveal a significant predominance of hits as the latter category. For example, on jstor.org, the search for ‘anthropology “big data”’ yielded 33,079 results; while searches for ‘anthropology “machine learning”’ and ‘anthropology “data science”’ yielded 206 and 148 results respectively.⁹ Many anthropologists are grappling with the implications of the relatively “big data” and the practices accompanied by it (such as Boellstorff 2015, Feldman 2017, Gray 2013, Lane 2016, Elish 2018, Norvaisas 2014, Wang 2013). This is not inherently bad, yet it betrays a lack of full understanding of the issue, particularly an on-the-ground, ethnographic view, even such a vantage point is the hallmark for discipline as a whole. As a theoretical construct focusing on big data magnifies the scale and supposed importance of the topic (implicitly amplifying its supposed *bigness*), which often makes the topic seem all-encompassing, hegemonic, and unified.

This obsession with big data defines the sets of intricate, fluid, and intermeshed practices as a single, totalizing system (or at least one attempting totalization), and through its emphasis on how that system influences the anthropologists’ lives (in most cases as a threat) further reifies these complex and multifaceted phenomena as a single totality. Instead of being a total system, big data, in conjunction with data science and machine learning, constitute complex sociotechnical phenomena, which definitely exists yet is pretty fluid/dynamic, decentralized (yet sometimes highly centralized), and complex. This airplane- or space-level view stifles an on-the-ground understanding of the topic developed through ethnography. Refocusing machine learning

⁹ I conducted these searches at 3:45 pm (U.S. Central Time) on April 25th, 2018. Note: In the search, I put quotes around the phrases *big data*, *data science*, and *machine learning* so that for each, it would return the whole phrase not, for example, search for all uses of the term “big” and “data” separately.

and on data scientists would shift anthropologists and other social analysts away from the totalizing threat of big data towards a more refined, nuanced, and intricate view of the topic.

The wider public has had a tendency to create all-encompassing myths or dramas about the recent algorithmic technologies called machine learning (c.f. Ziewitz 2016:5), and analysis of this topic by anthropologists and other social researchers has also invoked this myth-making. By *myths*, I mean moralizing narratives used to explain some phenomena in the world (a concept, institution or social group, a social force, metaphysical and/or physical entity, etc.), often used to explain, justify, and/or advocate for an idealized version of certain social roles, certain power structures, institutions, or other individual or social entities. There are two common narrative mythological depictions that surround data science, machine learning, and big data: viewing the algorithms as gods revolutionizing the world or as demons threatening our lives.

In the former, data science and machine learning promise a technological revolution that will help solve the world's problems and/or transform society, with the details on what those solutions may be, how effective they may be, and what exactly this societal transformation looking like varying depending on the teller. Peter Domingo's book, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (2015b), exemplifies that trend. Beyond a subtitle, which could serve as a succinct description of this myth in of itself, Domingo presents a historical narrative in which machine learning is transforming the world: "Society is changing, one machine learning algorithm at a time" (xiv). Here machine learning and the wider discipline of data science is a god, revolutionizing the world.¹⁰

¹⁰ This myth seems less prevalent within anthropological circles (or related social disciplines), and since this paper focuses the anthropological community, I will not focus on it. If I were writing this paper to a different set of audience, say machine learning or data science enthusiasts, I would develop the focus very differently. In anthropological contexts, this myth is primarily significant from the fact that the other myth, which anthropologists do tell, is a counter response to this myth.

In direct response to the deification of machine learning, many social analysts, including anthropologists, science and technology experts, algorithm studies scholars, etc., seek to render these forces into a demon. In this second myth, machine learning and other data science algorithms are just as powerfully singular and all-encompassing, but the supposed god is a really demon: a destructive entity threatening our lives. Both narratives reify the phenomena instead of understanding it/them in its multiplicity. The anthropological focus on big data arises out of this second myth and does not facilitate a questioning of that myth. Focusing on machine learning (one major type of on-the-ground practice in this topic) and data scientists (the major doers/workers on the ground) through ethnography would force anthropologists to slowly question this myth narrative and provide a more constructive view of the phenomena. Bringing the conversation away from an exclusive focus on big away has potential to bring new insights into the theoretical, sociocultural, and ethical implications of data science work, and ethnography is the best way to do this.

Discussing algorithms in general, Malte Ziewitz (2016) provides an interesting description of this process of myth-making within algorithm studies scholarship, which helps make sense of myth-making for machine learning algorithms. He describes an “algorithmic drama,” a seductive script, which “much recent talk about the role of algorithms in public life tends to follow,” including scholarship and academic conferences on algorithms (2016:5). This drama typically acts:

- Act 1: Introduce algorithms as powerful, consequential actors threatening human individuals, albeit subtle and flexible which makes them resistant to change (5)
- Act 2: Pick up on the complexity of algorithms and the difficulty analyzing them but are definitely a political concern (6).

This drama is appealing because it induces the algorithm “into a more or less ready-made system of conventional politics with its established cast of actors and long-standing concerns about agency, transparency, and normativity” (6). Chandler (2015), Eubanks (2017), Perrotta (2018), and Roark (2018) are all examples of works that recite versions of this drama. He situates this within the wider social tendency to view as either “straightforward tools or mythologies to be debunked” (Malte 2016:4) – each of which represent the two mythologies discussed above. The stable concept of an algorithm implied within fades in more detailed analysis “at the hands of ethnographers, historians, and sociologists” (8).

The most common telling of the myth of the dangers of machine learning and other data science algorithms typically center around telling horror stories of recently developed ethically questionable algorithms, such as the following:

1. Stanford-based study (Wang 2017) that sought to predict whether an individual was gay or straight based on facial features
2. Crime-prediction algorithms used to “predict” likely offenders of violent crime in various cities, such as one developed in London (c.f. Kelion 2014)
3. Google image search result, which misclassified black faces as gorillas (Noble 2017:5-7).

These are all ethically horrid instances negatively impacting a historically marginalized population. Social analysts typically invoke these examples to justify the regulation of algorithm development. These stories function as a mythological narrative depicting programmers and software developers involved in the development of these systems as naïve, incompetent, prejudiced, and/or malicious and justifying the specific role of social researchers to ethically regulate over them (c.f. Crawford 2017, as one potential example of this).

These algorithm creations and uses are definitely problematic, but simple retelling of fails to discuss the subtleties and intricacies of the complex terrain that is data science, in favor of simply exhibiting its most horrifying monsters, in a way analogous to the exhibitions of the supposedly strange creatures in other parts of the world. The closest parallel would be to describe anthropology and ethnography by providing case studies of the most ethically abusive work by anthropologists (such as certain high profile projects by certain anthropologists with the U.S. military or anthropological work used to justify and enable colonial oppression) in order to justify why social researchers should either not work data scientists as a whole or should require ethical oversight over them. This narrative fails to understand how these egregious project occurred within the wider context of the discipline, how those in the discipline understood and evaluated those projects, and how it influenced further ethical considerations by those within the discipline.

The work of Nick Seaver, Suzanne Thomas, Dawn Nafus, Jamie Sherman, and Taina Bucher provide a rich, theoretical alternative images for how to move beyond mythologizing big data: Nick Seaver through an elaborate ocean allegory and images of bastard academics (both bastardized algebra or mathematics and bastardized anthropology) and the others through concepts of machine learning algorithms as fetish. All of their visions entail using ethnography to understand data scientists and machine learning (and other) algorithms on the ground, through ethnography, as opposed to sky- (or maybe for Seaver, ocean-) level analysis of big data.

Nick Seaver in “Bastard Algebra” (2015a) focuses on the whole ecosystem(s) involved in these phenomena. Based upon a metaphorical images of deluges or floods of big data, he employs an elaborate ocean analogy to describe the different facets involved:

The ‘vast gulf’ between mathematical analyses and daily life... has been filled by a flood of data – credit card transactions, page views, the location of your cell phone, clicks, likes, and shares. The flood sustains new calculations that look less like bastard algebras and more like full-blooded formalisms, closer to the descriptive immersiveness of ethnography than the reductive genealogy of the family tree. As the collection of data quickens and sprawls, the floodwaters threaten to overtake the life they nominally represent: personalization algorithms produce a milieu for our activity online, shaped by and for the collection of data. (32-33)

These systems are made up of many distinct habitats:

1. The businessperson in the rowboat “looking for ‘insight’ and producing insight” by “making lures,” where “big data is used to attract and convince, to capture and compel rather than to prove” (2015a:33).
2. Life on the shore with “waves of blue 1s and 0s [crash] behind you,” where the “real” work of data science happens in all its boring, quotidian splendor, yet also where most creatures – that data scientists and work on algorithms – live¹¹ and in which the “frog’s eye view” or amphibious perspective in and out of the water are (41).

Even though both of these may require ethnography to understand, Seaver’s imperative is for anthropologists “to get off the boat and venture ashore,” which is a classically ethnographic thing to do (2015a:43). “From the shore, we can see the variety of people, epistemologies, and methods that constitute ‘data science,’ in addition to “the countless choices involved in cobbling

¹¹ From a purely ecological perspective, in the ocean, the most diverse ecosystems filled with the most densely packed variety of life (like coral reefs) typically occur in the shallow waters reasonably close to the shore, as opposed to in the open ocean. In addition to the liminal, “in-between” nature of the shore space, Seaver seems to implicitly capture this ecological diversity in his analogy as well.

big data together, moments of ambivalence and constraint, and the thickness of formalism in practice” (43). Anthropologists need to move beyond simply using ethnography to chart “a course back to the ‘real facts’ of relations” or “rehearse the old relationship between [qualitative and quantitative] methods,” we should use this both “to break down preconceptions about” data scientists and to examine how the methods relate” (43) (the latter I will discuss in more detail in Section 2).

Within Seaver’s analysis brings three important themes to the fore. First, through the image of big data as an ocean provides a compelling context, he demonstrates there are multiple subsystems, perspectives, and angles within (such as a fishing boat and shoreline). Not only do these need to be understood ethnographically in of themselves, but it demonstrates how anthropologists, through their focus on big data, analogously trying to look at the vast, seemingly all-encompassing ocean, have failed to understand the most significant sites: where this data runs up onto and shapes the shore. Second, this shore is the world of the data scientists and machine learning algorithms, and a shift towards machine learning and data science could reorient anthropologists to these sites. Finally, ethnography – or the “frog’s-eye view” – is the best way to understand what data scientists and machine learning algorithms are doing on the shore.

Suzanne Thomas, Dawn Nafus, and Jamie Sherman (2018) and Taina Bucher (2016) each advocate for understanding the current implications of faith in machine learning algorithms among developers. Thomas, Nafus and Sherman situate their analysis within a conception of machine learning algorithms as fetishes. Fetishes embody and distribute “power as capability, promise, faith, and possibility” based on attributing power from other entities (such as interpersonal relations, sociocultural/economic forces, etc.) (2018:4). This fetishization of machine learning algorithms positions them “in ways that make algorithms promise more than

they can deliver in strictly material terms.... This is the moment of social creativity when faith in a promise delivers possibility” (Thomas 2018:4).

As such, analyzing these algorithms as fetish allows anthropologists and other social scientists to both analyze the network of relations in which these algorithms exist, the ascription of meanings per their status in these relations, and how these computerized objects function as mediums within, without denying the creative potential such ascriptions of meaning entail. In contrast to the common simple accusation of something as a fetish, which “purports to unmask and yet reveals its own anxieties in the process” (2018:9), their work “testifies to the generative possibility of believing for a moment in the powers of algorithms, but only if we stop short of demonizing or deifying them” (10). Thus analyzing algorithms as fetishes provides an alternative framework to the mythological narratives of glory and terror discussed above. They do specifically by situating their focus on the sociocultural role of machine learning algorithms. Their emphasis is purely on meanings attributed to algorithms by human persons, which does not consider the ways the algorithms themselves in shaping and influencing their own role and ascriptions of meaning. Despite this, they provide a revitalizing reorientation for the discipline when other anthropologists have at times failed to look beyond the either supporting or unmasking these algorithms’ mythologies.

Likewise, for Bucher, understanding the fetishization of machine learning algorithms or imbuing of powers beyond their immediate scope is necessary to properly analyze their role. “The notion of fetish... allows us to recognize the boundaries and limits to knowing software in the first place” (2016:89). For her, algorithms are *talk-makers*, meaning that “what the algorithm does is not necessarily ‘in’ the algorithm as such” but in “the ways in which they are articulated, discussed, and contested in the public domain” (2016:90). “Focusing on the notion that

algorithms are black boxes grants them a special place in the world of unknowns that perhaps is not fully deserved” (2016:86). These algorithms are not simply a black box, but “a lot more gray, fluid, and entangled,” and researchers must analyze them as such (2016:94).

Implied within Bucher’s analysis is a view of machine learning algorithms as social entities. These systems have meaning and are able to perform (or fail to perform) based on sociocultural meanings imbued to them. Given this, ethnography is crucial to understanding how these systems function socially. Because of the social complexity and messiness of these algorithms, Bucher discourages any one recipe for understanding these systems (2016:85), but proposes a mode of inquiry called *technography*, which based from anthropological understandings of ethnography, seeks to describe and observe technology by examining “the interplay between a diverse set of actors (both human and non-human)” and “the norms and values that have been delegated and materialized in” these technologies” (2016:86). “While the ethnographer seeks to understand culture primarily through meanings attached to the world by people, the technographic inquiry starts by asking of what the software itself is suggestive” (2016:86).

I would suggest that this mode of inquiry is more than just “ethnographic-like” but actually is ethnography (albeit of a slightly different entity(s) than traditional, human-centered ethnographies) because her method focuses on how an agent (in this case, an algorithm) or set of agents embody, reflect, catalyze, and create meaning and knowledge in sociotechnical systems (c.f. Section 3, where I take up the notion of machine learning algorithms as agents in more detail). This is only a semantic distinction, not intended to detract from the brunt of her argument. She compellingly demonstrates the complexities involved in understanding how algorithms interrelate and exist within their lived worlds and that analyzing this fully requires a

myriad of approaches instead of a one-size-fits-all methodology (2016:91-93). The term *ethnography* can either encapsulate this methodological fluidity and need for multi-dimensionality (something it has done in other contexts) or simply provide one aspect in the overall methodological repertoire.

Anthropologists and other social researchers need to seek to understand the whole loose set of technological phenomena for which the terms *big data*, *machine learning*, and *data science* typically signify. To understand the whole system, we need to move away from a focus on big data towards machine learning and data science. Ethnography is definitely the best way to understand these second two sociotechnical entities, since it can in-depth picture of what they look like on the ground. This will help reformulate the single, totalizing mythological depictions of complex and knotty sociotechnical phenomena. Seaver's oceanic analogy and Thomas's, Nafus's, Sherman's and Bucher's theorizations of the machine learning algorithms as fetish provides a starting point for how to do that.

Section 2: Bridging the Gap between Ethnography and Machine Learning

Representatives of quantitative and qualitative research methodologies (the latter of which includes anthropologists) have long debated each other (c.f. Mertens 2018:23 for one example of this in a particular context). These academic "wars" have influenced the reception of machine learning systems among anthropologists (Seaver 2015a:43, Madsen 2018:31). Even though, as with any change, there is some continuity with the old, machine learning seems to entail divergences in quantitative analysis, both in the types of questions it seeks to address and the methodology it uses to address these questions (Mackenzie 2017:40, 80; Chandler 2015:841,

843; Ribes 2009:1; Perrotta 2018:13). These divergences provide an impetus for a new discussion between ethnography (and other qualitative approaches) and machine learning (and other quantitative approaches).

In this section, I will lay the groundwork for what this new discussion ought to look like. I will first discuss how machine learning algorithms embody new strategies within quantitative and mathematical modeling, both in terms of the questions raised about the data and the methodological tools used to answer these questions. These represent new points of intersection/commonality with ethnographic research leading to a new starting points in this discussion. Second, I will analyze the creative strategies anthropologists and other ethnographers have developed to either incorporate machine learning into ethnography and/or conduct ethnographic analysis alongside of it. These provide the seeds through which this a new discussion has been growing.

Although it would take a separate essay in of itself to flesh this out in detail, in the qualitative vs. quantitative debate, several dichotomies have become common to represent both sides: inductive vs deductive, open-ended vs close-ended, flexible vs prescriptive, ground-up vs top-down, subjective vs logical mathematical. In my own personal experience, this has often created an association, particularly among qualitative researchers but also among quantitative researchers, between mathematics and close-ended, top-down, procedural analysis (Plessis 2016:60-61 is one example of this). As someone with a background in both mathematics and anthropology, I have been uniquely frustrated at this association, which seems to fly in the face of my own experience as a practitioner in both anthropology and mathematics.

Machine learning breaks these simple molds. Many machine learning systems (particularly unsupervised learning techniques) are both inductive, flexible and mathematical.

Such techniques provide unique connections with ethnographic methods, with iterative qualitative coding in particular: both unsupervised machine learning and qualitative coding involve deciphering patterns through iterating over data in order to generate knowledge and insight. In a clear shift away from other statistical techniques (like a t-test) which seek to develop a single mathematical model to encapsulate and speak to a certain type of situation, machine learning models seek to build a unique model for the particular data or situation given:

Faced with the impracticality of an analytical or mathematically closed form solution to the problem of finding a function, machine learners typically seek ways of observing how different models traverse the data. They replace the exactitude and precision of mathematically-deduced closed-form solutions with algorithms that generate varying solutions. (Mackenzie 2017:95)

This key mathematical shift towards building the model from data parallels qualitative ethnographic methods. On the qualitative side, ethnography, with its focus on grounding social research in the particular context and allowing that context to determine the specific needs or practices employed, is far more similar to machine learning than any other type of quantitative approach.

Furthermore, data scientists are increasingly seeking to answer interprevist questions on understanding the world in which people operate. These questions come closer to ethnography in the types of questions they seek to answer than they do towards traditional quantitative methods. Both are both interprevist or ethnographic, situating themselves the same epistemological field of seeking to understand the ‘whys’ of cultural behavior (Curran 2013:63) This provides a new space of commonality and discussion between data scientists and anthropologists, which may help both find common ground while circumventing the barren no-man’s-land between warring

traditional qualitative and quantitative camps. Such a discussion also provides the potential to include the far richer fields of mathematics (such as topology, abstract algebra, and graph and network theory) that incorporate multidimensional and temporal thinking into the mix yet are inherently removed from conventional quantitative analysis.

Several scholars have developed strategies for using machine learning in ethnographies: Giaccardi's (2014) *thing ethnography*, Geiger's and Ribes's (2011) *trace ethnography*, and Haines's (2017) *multi-dimensional ethnography*, and Madsen's (2018) *transversal collaboration*. These provide examples of how that conversation can go. Giaccardi emphasizes the use of machine learning in the process of analyzing things ethnography; while Geiger, Ribes, and Haines analyze how to use ethnography to understand virtual and data systems. Madsen explores ways to situate cross-collaborative work between anthropologists and data scientists.

Elisa Giaccardi, in collaboration with Chris Speed and a few other scholars, develops the concept of *thing ethnography* to articulate a way to incorporate the perspective of things through machine learning of data outputted by/for these things. The foundation of this is a dual contribution between things and humans: "As humans, we shape objects, and objects shape us in return" (2014b:377). "Equipped with software and sensors, things can begin to provide access to fields and perspectives that would not be otherwise attainable from human ethnographers", the collection and interpretation of which she calls *thing ethnography* (2014b:378). Machine learning is a major means by which to conceptualize and articulate these perspectives of things ethnographically, which is then used to further innovate: her research project "uses a combination of field studies, object instrumentation and machine learning to listen to what 'things' have to tell about their use, reuse, and deviant purposes, and it harvests this data to

inspire idea generation, fabrication, rapid prototyping, and business development generation” (2014a:1).

Her concept fits within the wider Internet of Things framework, basing the specific usage of machine learning for data-generating objects. This is a pretty specific type of object and type of machine learning – objects interconnected through a networked Internet of Things platform continuously generating data as such and the specific machine learning systems used to analyze the data from these objects – but she provides theoretical considerations applicable beyond this immediate scope. She describes how things possess their own agency and rhythm, not typically understood by exclusively human-centered perspectives (383-386), and how the specific things she analyzed ethnographically exhibit unique yet important movement, temporality, and agency (378).

Simply providing machine learning as the mechanism for articulating or making sense of this voice seems theoretically naïve, since these algorithms and the processes involved in gathering “data” from these things that these algorithms analyze are not simple reflections of the things themselves, but are *themselves* also clearly non-neutral social constructs that have their own force or agency to them. Her attempt to move beyond non-anthropocentric ethnographic methodology to understand the data worlds and to specifically use machine learning to do so is commendable. As I will discuss in Section 3, more anthropological work that explores how these non-human agents participates is in order.

Stuart Geiger and David Ribes discuss how to ethnographically analyze large scale digital infrastructure and organizations, and he has coined the term *trace ethnography* to describe a strategy to integrate social and computational methodologies in order to analyze these sociotechnical environments (Geiger & Ribes 2011:1). *Trace ethnography* analyzes the

documentary traces in these systems (such as “transaction logs, version histories, institutional records, conversation transcripts, and source code”) in order to glean “rich qualitative insight into the interactions of users” (Geiger & Ribes 2011:1). The ultimate aim is “to retroactively reconstruct specific actions at a fine level of granularity” through the assemblage “into rich narratives of interaction” (Geiger & Ribes 2011:1). Trace ethnography is especially beneficial in fast-paced and decentralized distributed and open-source software communities (such as Wikipedia, his primary research interest at the time) (Geiger & Ribes 2011:2-3), which could include many data science groups, communities, and networks.¹² In addition, it could help analyze the interaction of many computerized knowledge formation and recommendation systems (many of which employ machine learning techniques) as they converse and interact with humans digitally: shown by Gieger’s previous work analyzing bot- and human- interactions in Wikipedia (Geiger 2011:78; Geiger & Ribes 2011:2).

Trace ethnography has some clear limitations, however, the most pertinent of which is that it “does not allow immediately allow researchers to grasp the large sociocultural significance or history of the activities at hand,” yet its combination with other qualitative and quantitative approaches (Geiger & Ribes 2011:9). Despite such limitations, it is a good starting point for discussions of how to integrate ethnographic analysis with large scale data science algorithmic, given that such systems utilize these computerized documentary/data traces ripe for analysis. Trace ethnography easily complements large-scale quantitative analysis into the documents and (if possible and if it exists) traditional ethnographic analysis into non-virtual environment in which these traces were formed. Most importantly, trace ethnography takes seriously the

¹² For an example of this, see Matthew McCarthy (2017) for his continuation of Gieger’s work in Wikipedia to understand the semantic web within on Github. (Github is an open source frequented by data scientists and other programmers to share work and collaborate.)

complex and intricate existent of sociotechnical systems and seeks to analyze these ethnographically.

Julia Haines's concept of *multi-dimensional ethnography* takes trace ethnography as its starting point, seeking to incorporate its strengths yet provide methodological complements to patch up its potential holes (2017:128-130). Utilizing the concept of traces, she advocates for a multi-dimensional approach to analyze multiplicity of selves utilized in digitalized technology: "the boundaries and seams, the flows and the assemblages, and the multiple identities, agencies, and temporalities inherent in the use of digital technologies" (2017:130). She seeks to provide a holistic perspective that captures these multiple dimensions:

My aim was to get a more holistic picture of people's lives. And beyond the complexities of traditional ethnographic praxis, the biggest challenges were to understand the boundaries and seams, the flows and the assemblages, and the multiple identities, agencies, and temporalities inherent in the use of digital technologies. We humans do not now, nor have we ever, had just one self. But with digital technologies, we have an incredible multiplicity of selves -- selves that can exist and interact and shape things, even when we are not present. And as we continue to move into a machine-learned world, this will become an even more important force playing a role in our lives. These experiences are proliferating and add ever-increasing complexity to our lives. As researchers, we must adjust our research approaches, methods, and ways of sharing insights. We need grounded methodological and analytical tools to interpret things in a tractable way.

This builds on the type of documentary ethnographic work but includes a wide variety of other qualitative and quantitative data, methodology, and instruments (2017:131). She sees

Gieger's and Ribes's concept of trace ethnography as an important attempt to understand data streams ethnographically and to capture the digital dimensions of human selves present there, yet which fails to connect these traces to the wider (non-virtual and virtual worlds and dimensions of self that inhabit these) (2017:130). Her concept of *multi-dimensional ethnography* is her "humble start at developing an approach aimed to probe this multi-layered, multi-faceted realm of exploration, laying the foundations for future investigations" (2017:131).

Her theoretical vision lays the foundation for the direction I think ethnographies of these computerized algorithmic systems need to go. She has bitten off more than she can chew, but that is because she has marked off a large mountain of food, which will take time and personnel to consume.

Mette My Madsen, collaborating with Anders Blok and Morten Pedersen, develops a concept called *transversal collaboration* to describe the "instantiating forms of non-coherent, intermittent, yet productively mutual co-shaping among partially, connected knowledge practices and practitioners" evident in a joint project between data scientists and herself. The word *transversal* allows us to talk about intersections between these two areas without conceiving of either as distinct, separate entities (2018:6-7), in a way that recalls for me topological discussions of mutually overlapped but necessarily intersecting spaces.¹³ Her role as an anthropologists involved "oscillation between practising ethnography *in* a computational social science framework and doing an ethnography *of* the different data practices and infrastructures involved" (2018:5-6, italics in original) – which I have navigated in my own work on similar types of

¹³ A simple example of this could be the set of all rational numbers between 0 and 1 (exclusive), and the set of irrational numbers between 0 and 1 (exclusive). These two sets overlap or encapsulate each other infinitely (between any two rational numbers between 0 and 1 is an irrational number and vice versa) but have no overlap with each other. This is a simple example of more complex topological spaces, which may surround yet intersect each other in unique and unexpected ways.

projects. As opposed to the distant gaze of STS observer (2018:2), she employs the term *transversal* to her attempts to simultaneously find mutually co-productive intersections and to work up “the line” of power asymmetries (2018:6, 31). She intends *transversal collaboration* as a new “conceptual vocabulary and analytical language” to situate this complex process of coproduction and critical examination (2018:6). This has allowed her to see her role “as a subject to a line of flight that criss-crosses and stitches together otherwise disparate data worlds” (2018:7). This is necessary for both understanding and addressing “new social data ‘complementaries’ and their epistemological, ethical, and political ramifications” (2018:2).

She argues that anthropologists working in this area must be willing “to transpose *both* ethnography *and* computational social science on to a *third* plane of social investigation, one that is transversal to but not co-terminus with either” (2018:32). The ultimate goal in describing her work is that “the transversal line of flight across otherwise disparate data worlds [in her work]... may come to inspire more experimental endeavours at the mutually intersecting, yet still too often still mutually hostile endeavours of ethnographic and ‘big’ social data” (2018:32). Even if, she heightens and/or essentializes the supposed disparities between anthropology and data science, her concept of *transversal collaboration* provides a theoretical framework for the unique types of subjects, activities, and products created by intersecting these two fields. The perspectives of anthropologists and designers discussed below exemplifies this potential space.

Section 3: Machine Learning Algorithm as Agents

Machine learning algorithms provide one major context in which to broaden our understanding of what has agency and what participates in a culture, given that they develop

their own “knowledge” outside of the immediate domain of the human creators. In this section, I will discuss the theoretical implications of the potential agency of machine learning algorithms. By agency, I mean the ability to influence other entities, which I view as synonymous with being in relationship with. I will first demonstrate how this discussion fits within the wider theoretical discussion in anthropology over agency, particularly by its unique intersection of Foucaultian and Latourian concepts, and then discuss use Adrian Mackenzie’s recent book on machine learning as one major example of how to conceptualize the agency of machine learning algorithms.

My discussion on the anthropology of machine learning situates within the wider debate anthropological discussion over agency. *Agency* has clearly been a key term within anthropology, and certain theorists within the last several decades have raised questions over what has agency, particularly of non-human agency. For example, Bruno Latour and Webb Keane each discuss the narrative of modernity hinges on regulating agency only to humans, specifically away from the material world. In *We Have Never Been Modern*, Latour focuses on the tendency by moderns to dispense agency only to humans (1993:10-12, 128). Following Latour, Keane explores how modern Europeans sought to discard the agency non-human entities, particularly language, words, and sounds, by denying that they possess any agentic influence (2007:23-25, 47-49, 55, 223-224). This connects to the wider question, do non-human entities have agency, and if so, can they participate in a culture?

When anthropologists and related scholars theorize about the non-human agents, two common broad types of non-human agents are material entities, and social processes/structures. Latour provides a major theoretical foundation within anthropology for the former and Foucault for the latter.

Latour's work wrestles significantly with the agency of material entities (objects, natural forces, etc.), particularly in his analysis of natural science. In *We Have Never Been Modern*, he problematizes the distinction between society and nature, and the distinction's implied separation of human and natural agencies into completely disparate natures (1993:3-7). His actor-network theory (AT) – in conjunction with several other theorists – further provides an alternative way of conceptualizing agency (my own definition of agency has been shaped by Latour's concept of an actor), which incorporates both human and natural actors into an overall system (2017, 2). AT helps develop a unique theory of study which seeks to analyze the interaction of different actors/agents that are often theorized as existing in different categorically distinct plans:

Three resources have been developed over the ages to deal with agencies. The first one is to attribute to them naturality and to link them with nature. The second one is to grant them sociality and to tie them with the social fabric. The third one is to consider them as a semiotic construction and to relate agency with the building of meaning. The originality of science studies comes from the impossibility of clearly differentiating those three resources. (2017:1)

Foucault analyzes the way in which various institutional bodies (prisons, psychiatric wards, clinics, practice of therapy, etc.) shape individuals through discipline. Overtime, he establishes an elaborate understanding of *power* to analyze how these institutional bodies have helped form and shift bodily habits, persons and subjects, and the formation of knowledge (c.f. Foucault 1972:51-52, 233-236). *Agency*, the ability to affect or influence, represents the flip side of the conceptual coin to Foucault's power: power focuses on what is influenced and agency on what influences. In some of his analysis, these institutional powers and forces seem to possess a wholly or primarily destructive or harmful influence over individuals (e.g. Foucault 1965, 1973,

and 1977a), but in later his works, Foucault clarifies that institutional power can have both positive and negative aspects: both through its ability to catalyze the production of knowledge (1977b:51-52, 93) and through its formative role over human selves (1988:4-7).¹⁴ Either way, Foucault provides one major theoretical launching point for anthropologists to conceptualize the potential agency of institutional bodies.

Machine learning algorithms straddle both. They are both components of objects that we interact with (like a phone or a website), but they are also a process that humans and computers create and undergo. This complicates typical strategies of classifying agency and complicates a simple explanation of a machine learning algorithm's potential agency.

At times, machine learning algorithms can seem like objects: they are clear entities interacted with by other entities (both human and non-human), albeit often virtual or computerized. Such algorithms frequently provide a core component of objects that a human may interact with: a phone one talks to, a car that drives, or a website newsfeed that provides links to look at. More technically, machine learning algorithms operate within other computational systems, interacting with these to develop an overall entity or result for the human, in which they may provide an exclusively backstage role or possess a front-facing component. Thus, they are clear objects within these computerized systems and can often appear as literal objects in everyday life.

¹⁴ I see this emphasis on the harmful and violent nature of the creation of individual persons and subjects to be representative of the modernist tendency to see non-human agents (in this case social institutions) as dangerous out of fear that they may potentially limit the agency of human individuals. Illustrating this thread within Foucault's work and representations of Foucault's work by anthropologists and demonstrating how he also counters this tendency is a fascinating and much needed research, yet it is beyond the scope of this paper. At some point in the future, I may try to analyze this directly.

At the same time, they also both possess an institutional presence and arguably are institutions processes or technologies themselves. As part of a social world, machine learning algorithms can influence social bodies and institutional often specifically in ways that recall a Foucaultian understanding of formation of selves through discipline, illustrated by my example below. They also do this in how they interact and help form the software systems they are a part of. At the same time, as Adrian Mackenzie discusses (more details below), machine learning algorithms internally are their own disciplinary regime or technology: they involve iterating through data, making data through a prescribed set of criteria. As such, they are both products of and help form, themselves, institutional agents through discipline. The fact that they represent both an object and an institution is not surprising considering the fact that theoretically, all entities (including humans) are both made up of entities within for which it exhibits forms of agency reflective of depictions of institutional power and external entities with which it interacts as like an object, but considering that machine learning algorithms specifically blur this distinction within the technological world, they may be a useful basis for blurring those categories.

Consider the following machine learning algorithm as an example of potential ways in which a machine learning is an agent and a social entity. The plenary speaker at a data science conference I attended, who was on the executive team of a data science consulting company, told the following story about a project his company worked on. The project was with a large plantation in Bangladesh (according to him about the size of New Jersey) specializing in a certain type of wood. The plantation had a team of about fifty botanists whose job was to check the trees and ensure that they were healthy. The company developed a machine learning algorithm to predict which trees would be the most likely become ill and when, and from this,

developed a system which recommended which trees required an immediate assessment, which needed a secondary assessment, and so on. This changed the nature of the work for the botanists and the nature of the relationship between these botanists and their management structure which hired them.¹⁵

In this story, the machine learning algorithm had a very clear social agent. It provided recommendations to botanists about an action to take: which trees to go check. This relationship clearly functioned within a series of social and institutional relationships not only between botanists and managers but also between the data scientists who built the algorithm and both these groups and between the two companies as a whole. In this case, the algorithm is also subsuming a function typically associated with a human: determining the “best” trees to monitor first and providing recommendations to the user accordingly. It is both a computerized object that humans interact with, and could potentially play a role in shifting the institutional quotidian disciplines of the botanists. Now just because it communicates with humans – even fulfilling a role in that communication process that typically a human would fulfill – does not mean its agency is the same as that of a human. Like any entity in a social system, it both influences the dynamics within that system and is in turn influenced by them in unique ways. This ability to influence these dynamics provide the basis for why my claim that machine learning algorithms possess agency, and the fact that it influences these in a sociocultural system (like the institutional body of a company/plantation) makes this algorithm a participant in the

¹⁵ The speaker’s moral and purpose of telling this story was to demonstrate that data scientists, when developing these algorithms, need to better consider the social systems in/for which these technologies are made, specifically the perspectives of on-the-ground implementers of that job function at the company (as opposed to business managers who their formal interactions with the company are typically primarily with). I could unpack a lot within this story about the power dynamics and strategies for design, which that is the topic for another paper. Exploring this in more detail in future ethnographic inquiries analyzing this algorithm or similar types of algorithms would be a worthwhile endeavor.

sociocultures it is both developed in and utilized in. The fact that anthropologists do not have a readily apparent language or theory for what it means for non-human entities (either created technologies like algorithms or otherwise) participate in sociocultures is a troubling, and because of certain characteristics, machine learning algorithms can become a good springboard upon which to establish one.

In addition, the role the machine learning plays or what it “does” is definitely not reducible to the human entities (human individuals and groups) involved. This is the case for any technology, but because of the way, machine learning systems generate insights iteratively that are typically beyond the immediate scope of their creators, this can be extremely apparent. In the botany algorithm discussed above, the data scientists who built it developed a set of guidelines for the program, from which it would build a specific model for how to determine which trees are healthy based on inputting relevant data about the trees. The programmers themselves do not know what the algorithm will predict, nor even what exact measures it will use to do that, only the criteria used to develop the measures used to determine the result, which in turn become shaped overtime by the data fed. Most such algorithms, and I have no reason to think this one is any different, shifts its measures to incorporate that data, making even determining what the model says a moving target.¹⁶

In addition, the “knowledge” they learn or develop is not always understandable by humans. Front-facing machine learning algorithms, which are intended to create something directly seen and/or interacted with by a human user (as opposed to back-facing machine

¹⁶ This is one problem about the use of the black-box analogy to describe machine learning algorithms. More traditional algorithms given one output for anything inputted, so even if you did not know the contents of the inside (which is not always the case for algorithms), then you could learn about by testing what outputs it presented for a set of inputs. Many machine learning algorithms will shift (albeit not consistently) based on the data encountered so that if you put an input it, it will incorporate it and change what it does accordingly.

learning algorithms, which interact with other algorithms exclusively), often intentionally create a human-intelligible output (such as a recommendation), but even that is only related to, not the same as, what the algorithm knows: only a synthesis of its “learning” into a specific context (a specific instance, usually with a specific criteria). Processes for determining how it is recommending or what criteria it is using is difficult to relate in a way intelligible for a humans.¹⁷

In his book, *Machine Learners: Archaeology of a Data Practice*, Adrian Mackenzie (2017) has conducted one of the most systematic analyses into the potential agency of the machine learning process. For him, the primary agent in the machine learning process is a human-computer relationship, created by machine learning processes. Out of this relationship, new human and computer subjectivities, identities, and relationships emerge. His concept of “machine learners” (the primary entity in this relationship) “refers to both humans and machines or human-machine relations” (2017:6). Drawing from Foucault, he argues “that data practices associated with machine learning delimit a *positivity* of knowing,” that is a space for the construction of (certain) knowledge (6, italics in original), based on a new “operational formulation” (211). “The formal, mathematical abstractions, and certain transformations in the production of software associated with machine learning in tandem as diagrammatic processes that organize, and assemble human-machine relations” weave and produce this positivity (211). The formation of mathematical thinking into algorithmic processes, which he terms as a diagrammatic movement, enables or produces many possibilities, including the statistical transformation through algorithmic diagrams (23), vectorization of data into a mutable and expandable entity (69, 73-74), a reconfiguration of what counts as pattern (126), the development

¹⁷ I am including humans with a wide range of technical expertise. Humans with the technical expertise to understand the intricacies of the process mathematically and computationally, even the writers of the code, often do not know how to intelligibly understand the knowledge the algorithm forms.

of optimization as a form of observation (96, 210), probabilization of knowledge (105), a shift from what is visible to what is operational (149), and the assigning of new subject positions (180).

This diagrammatic movement, which parallels of the type of discipline of knowledge Foucault analyzes for institutions, is the primary agent or driver of influence in Mackenzie's account. Machine learning algorithms are "intersocial" "in the sense that they bring together different mathematical, algorithmic, operational, and observational processes," which "can constitute a 'new model truth' and can unmake 'preceding realities and significations'" (2017:101). Throughout the book, he emphasizes the agency or role machine learning processes play in developing a certain type of interface both among the various nonhuman agents within, like mathematical models, data, variables, etc. and between these entities and humans, out of which new subject positions and knowledge emerges (180). Machine learning techniques assign these, at times, in seemingly distinct but combinable ways: hierarchically yet mobilized (186) and without a centralized subject yet using the control of hyperparameters to minimize error to organize the process (198-200). These institutional techniques influence both nonhuman and human entities in shockingly parallel ways: for example, Kaggle competitions – a major forum for data scientists to come together to build machine learning algorithms – possess, for him, a very similar set of institutional techniques as the machine learning algorithms themselves (201-206).

Mackenzie's elaborate concept of machine learning provides the one major basis for describing the potential agency of machine learning algorithms. Drawing primarily from Foucault, he discusses the agency of machine learning systems through its institutional potential to configure and regulate the intersocial relationships between several entities within, both

human and nonhuman. The latter entities he defines as machine learners, which become formed into unique subject positions through the creation of these institutional techniques (or technologies). Out of this emerges a new formation of knowledge. He draws pretty heavily from Foucaultian conceptualizations of agency as primarily reflected in institutional control, and thus, he deemphasizes the potential machine learning algorithms as objects interacted with by things external to it. Viewing how these agents interact outside of themselves is necessary for understanding them as full participants in sociocultures and as more than simply a hegemonic institutional force, whether positive, negative, or both. Yet, these institutional processes latent within machine learning techniques *do* have a significant agentic influence, which cannot be ignored, so his emphasis on this by no means misplaced or erroneous.

Most importantly and foundationally, however, the social world in which machine learning techniques operate encapsulate both human and nonhuman agents, including the data, mathematical abstractions, variables, etc. This exemplifies a necessary trend in anthropological trend in this topic away from an exclusively human-centered view of the process. Published in November of 2017, this is the first major book by an anthropologist to tackle machine learning as a practice directly, and as such, it could prove a decisive theoretical starting point for anthropological analysis of machine learning.

Machine learning systems seem to have agency and participate in cultures. We need to reconceptualize our understanding of both to understand how a computer program could have agency. They also blur implied distinctions between objects and institutions, an already blurry concept. These distinctions of things need to be redefined. As shown by the Mackenzie's book, this is a complicated but worthwhile endeavor in order to understand how both machine learning algorithms and other nonhuman entities participate in sociocultures.

Conclusion

Anthropologists should explore machine learning anew in order to revitalize their understanding of the interconnected sociotechnical phenomena of machine learning, data science, and big data, and I use the phrase “anthropology of machine learning” to denote this potential future endeavor. An ethnographic focus on machine learning and data scientists – necessitating a conscious shift away from big data – would facilitate this. This would help foster new connections between anthropology and data science and within the qualitative/quantitative battlefield, this could help generate new connections with a newly rising perspective more potentially amicable to ethnography and other anthropological methods and modes of thinking. Finally, ethnographically understanding machine learning algorithms could benefit the broader theoretical discussion within anthropology of what agency is and what it means to be agentic, because of the rigorous yet important agentic nature.

Bibliography

Abe, A., & Hayashi, M. (2016). On communication assistance via bots —towards IMDJ. *ScienceDirect*, 1657-1665.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 734-749.

Anderson, R., & Sharrock, W. (2013). Ethical Algorithms: A brief comment on an extensive muddle.

Anderson, S., & Anderson, M. (2009). *How Machines Can Advance Ethics*. Retrieved from Philosophy Now:

https://philosophynow.org/issues/72/How_Machines_Can_Advance_Ethics

Bail, C. (2014). The cultural environment: measuring culture with big data. *Social Theory*.

Baumer, E., Mimno, D., Guha, S., Quan, E., & Gay, G. (2017). Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence?11. *Journal of the Association for INformation Science and Technology*, 1387-1410.

Bharwani, S. (2006). Understanding Complex Behavior and Decision Making Using Ethnographic Knowledge Elicitation Tools (KnETs). *Social Science Computer Review*, 78-105.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Bleecker, J. (2005). Why Things Matter: A Manifesto for Networked Objects — Cohabiting with Pigeons, Aphids and Aibos in the Internet of Things. *Creative Commons*.

Blue, A. (2017, 8 28). *Digital Archaeology Project to Use Big Data*. Retrieved from University of Arizona News: <https://uanews.arizona.edu/story/digital-archaeology-project-use-big-data>

Boellstorff, T., & Maurer, B. (2015). *Data, Now Bigger and Better!* Chicago: Prickly Paradigm Press.

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication, & Society*, 662-679.

Brondo, K., & Bennett, L. (2012). Career Subjectivities in U.S. Anthropology: Gender, Practice, and Resistance. *American Anthropologist*, 598–610.

Bucher, T. (2016). Neither Black Nor Box: Ways of Knowing Algorithms. In S. Kubitschko, & A. Kaun, *Innovative Methods in Media and Communication Research* (pp. 81-98). Cham: Palgrave Macmillan.

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*.

Burri, T. (2017). Machine Learning and the Law: Five Theses. 1-4.

Burrsettles. (2012, 2 5). *Machine Learning and Social Science: Taking The Best of Both Worlds*. Retrieved from <https://slackprop.wordpress.com/2013/02/05/machine-learning-and-social-science/>

Caplan, R., & boyd, d. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data & Society*.

- Cardon, D. (2016). Deconstructing the algorithm: four types of digital information calculations. In R. Seyfert, & J. Robarge, *Algorithmic Cultures: Essays on Meaning, Performance, and New Technologies* (pp. 95-110). London: Routledge.
- CGP Grey. (2017, 12 18). *How Machines *Really* Learn. [Footnote]*. Retrieved from
<https://www.youtube.com/watch?v=wvWpdrfoEv0>
- CGP Grey. (2017, 12 18). *How Machines Learn*. Retrieved from
<https://www.youtube.com/watch?v=R9OHn5ZF4Uo>
- Chandler, D. (2015). A World without Causation: Big Data and the Coming of Age of Posthumanism. *Millennium: Journal of International Studies*.
- Chatfield, T. (2017, 9 20). *AI will simplify talent acquisition*. Retrieved from
<https://venturebeat.com/2017/09/20/ai-will-simplify-talent-acquisition/>
- Cheney-Lippold, J. (2017). *We Are Data: Algorithms and the Making of Our Digitalized Selves*. New York: New York University Press.
- Clement, M., & Guitton, M. (2015). Interacting with bots online: Users' reactions to actions of automated programs in Wikipedia. *Computers in Human Behavior*, 66-75.
- Countee, A. (2016, 7 31). An Engineering Anthropologist: Why tech companies need to hire software developers with ethnographic skills. *Ethnography Matters*. Retrieved from
<https://anthrocode.com/2016/07/31/an-engineering-anthropologist-why-tech-companies-need-to-hire-software-developers-with-ethnographic-skills/>
- Crawford, K. (2014). Big Data Anxieties: From Squeaky Dolphin to Normcore. *Epic*.

Crawford, K. (2017, 12 10). *The Trouble with Bias*. Retrieved from NIPS2017:

https://www.youtube.com/watch?v=fMym_BKWQzk

Croll, A. (2014, 2 4). *Data Everywhere: Data Anthropology, Quantified Self, Machine Data, Human Centered Design, and more*. Retrieved from <http://www.oreilly.com/pub/e/2997>

Cunningham, S. J. (1996). Machine learning applications in anthropology: automated discovery over kinship structures. *Computers and the Humanities*.

Curran, J. (2013). Big Data or ‘Big Ethnographic Data’? Positioning Big Data within the Ethnographic Space. *EPIC*.

Cyborg Anthropology. (2011). *What is Cyborg Anthropology?* Retrieved from http://cyborganthropology.com/What_is_Cyborg_Anthropology%3F

Deleuze, G. (1992). Postscript on the Societies of Control. *The MIT Press*, 3-7.

Diakopoulos, N. (2013, 8 2). *Sex, Violence, and Autocomplete Algorithms*. Retrieved from Slate: http://www.slate.com/articles/technology/future_tense/2013/08/words_banned_from_bin_g_and_google_s_autocomplete_algorithms.html

Domingos, P. (2012). A Few Useful Things to Know about Machine Learning. *ACM*.

Domingos, P. (2015). The Five Tribes of Machine Learning And What You Can Take from Each. *ACM*.

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books.

- Driscoll, K., & Walker, S. (2014). Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data. *International Journal of Communication*, 1745-1764.
- Ducheneaut, N., Yee, N., & Bellotti, V. (2010). The Best of Both (Virtual) Worlds: Using Ethnography and Computational Tools to Study Online Behavior. *EPIC*, 136-148.
- Duda, R., Hart, P., & Stork, D. (2000). *Pattern Classification*. Wiley-Interscience.
- Edwards, C., Beattie, A., Edwards, A., & Spence, P. (2016). Differences in perceptions of communication quality between a Twitterbot and human agent for information seeking and learning. *Computers in Human Behavior*, 666-671.
- Edwards, C., Edwards, A., Spence, P., & Shelton, A. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior*, 372-376.
- Elish, M., & boyd, d. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, 57-80.
- Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., & Rickman, A. (2016). First I “like” it, then I hide it: Folk Theories of Social Feeds. *Curation and Algorithms*.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- Evans, B. (2016). Paco—Applying Computational Methods to Scale Qualitative. *EPIC*, 348-368.
- Faßler, M. (2013). Human-Computer-Inter-Creativity: A Co-Evolutionary Approach.
- Feldman, J. (2017). Big data and ethnology. *Anthropology Today*.

- Fischer, M. (2009). *Anthropological Futures*. Durham: Duke University Press.
- Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge: Cambridge University Press.
- Foucault, M. (1965). *Madness and Civilization*. New York: Random House, Inc.
- Foucault, M. (1977). *Discipline & Punish*. New York: Random House, Inc.
- Foucault, M. (1977). *Power/Knowledge*. New York: Random House, Inc.
- Foucault, M. (1988). *The Final Foucault*. Cambridge: MIT Press.
- Friedman, U. (2012). Big Data: Anthropology of an Idea. *Foreign Policy*.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*.
- Geiger, S. (2009). Does Habermas Understand the Internet? The Algorithmic Construction of the Blogosphere/Public Sphere. *Gnovis: A Journal of Communication, Culture, and Technology*.
- Geiger, S. (2011). The Lives of Bots. In G. Lovink, & N. Tkacz, *A Critical Point of View: A Wikipedia Reader* (pp. 78-93). Amsterdam: Network Cultures.
- Geiger, S. (2017, 9 14). Computational Ethnography and the Ethnography of Computation.
- Geiger, S., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices.
- Giaccardi, E., Speed, C., & Rubens, N. (2014). Things Making Things: An Ethnography of the Impossible.

Giaccardi, E., Speed, C., Cila, N., & Caldwell, M. (2016). Thing Ethnography: Doing Design Research with Non-Humans. *DIS*.

Gillespie, T., & Seaver, N. (2016, 12 15). *Critical Algorithm Studies: a Reading List*. Retrieved from <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>

Gray, M. (2013). Big Data, Ethical Futures. *Anthropology News*.

Gray, M., Suri, S., Ali, S. S., & Kulkarni, D. (2016). The Crowd is a Collaborative Network. *Social and Behavioral Sciences: Sociology*.

Gustafon, S. (2016, 6 20). *The human-side of artificial intelligence and machine learning*. Retrieved from <http://ethnographymatters.net/blog/2016/06/20/the-human-side-of-artificial-intelligence-and-machine-learning/>

Haines, J. (2017). Towards Multi-Dimensional Ethnography. *EPIC*.

Hale, S. (2017). What is Data Science? *SAGE Publications*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hill, K., & Surya, M. (2018, 2 7). *The House That Spied on Me*. Retrieved from <https://gizmodo.com/the-house-that-spied-on-me-1822429852?rev=1518027891546>

Introna, L. (2016). The algorithmic choreography of the impressionable subject. In R. Seyfert, & J. Roberge, *Algorithmic Cultures: Essays on Meaning Performance and New Technologies* (pp. 25-51). London: Routledge.

Irani, L. (2015, 1 15). *JUSTICE FOR “DATA JANITORS”*. Retrieved from <http://www.publicbooks.org/justice-for-data-janitors/>

Jasper, A. (2004). 'I am not a goth!': The Unspoken Morale of Authenticity within the Dutch Gothic Subculture. *Ethnofoor*, 90-115.

Johnson, J. (1988). Mixing Humans and Nonhumans Together: The Sociology of a Door-Closer*. *Social Problems*.

Kaghan, W., & Bowker, G. (n.d.). Out of machine age?: complexity, sociotechnical systems and actor network theory. *Journal of Engineering and Technology Development*.

Keane, W. (2007). *Christian Moderns*. Berkeley: University of California Press.

Kelion, L. (2014, October 29). *London police trial gang violence 'predicting' software*. Retrieved from <http://www.bbc.com/news/technology-29824854>

Kelkar, S. (2014, June 18). *On the Porous Boundaries of Computer Science*. Retrieved from The CASTAC Blog: <http://blog.castac.org/2014/06/on-the-porous-boundaries-of-computer-science/>

Kelty, C. (2016). Two Fables. *Pulse*, 490-514.

Kippen, J. (1988). On the Uses of Computers in Anthropological Research. *Current Anthropology*, 317-320.

Kippen, J., & Bel, B. (1989). Can a computer help resolve the problem of ethnographic description? *Anthropological Quarterly*, 131-144.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Infrastructures & Their Consequences*. Los Angeles: SAGE.

Kockelman, P. (2013). The anthropology of an equation. *HAU: Journal of Ethnographic Theory*.

Kononenko, I., & Kukar, M. (2007). *Machine Learning and Data Mining*. Elsevier: Philadelphia.

Kubitschko, S., & Kaun, A. (2016). *Innovative Methods in Media and Communication Research*. Cham: Palgrave Macmillan.

Ladner, S. (2017, 12 7). Retrieved from WHY MACHINE LEARNING ISN'T ABOUT MACHINES: www.samladner.com/why-machine-learning-isnt-about-machines/

Lalwani, M. (2016, 8 16). *The next wave of AI is rooted in human culture and history*. Retrieved from <https://www.engadget.com/2016/08/16/the-next-wave-of-ai-is-rooted-in-human-culture-and-history>

Lane, J. (2016). Big Data and Anthropology.

Latour, B. (1993). *We Have Never Been Modern*. Cambridge: Harvard University Press.

Mackenzie, A. (2006). The Strange Meshing of Impersonal and Personal Forces in Technological Action. *Culture, theory & critique*, 197-212.

Mackenzie, A. (2012). More parts than elements: how databases multiply. *Society and Space*.

Mackenzie, A. (2012). Set. In C. Lury, & N. Wakeford, *Inventive Methods: The Happening of the Social* (pp. 219-231). London: Routledge.

Mackenzie, A. (2013). Programming subjects in the regime of anticipation: Software studies and subjectivity. *Subjectivity*.

Mackenzie, A. (2014). Multiplying numbers differently: an epidemiology of contagious convolution. *Distinktion: Journal of Social Theory*.

- Mackenzie, A. (2015). The production of prediction: What does machine learning want? *European Journal of Cultural Studies*.
- Mackenzie, A. (2017). *Machine Learners: Archaeology of a Data Practice*. Cambridge: The MIT Press.
- Mackenzie, A., & McNally, R. (2013). Living Multiples: How Large-scale Scientific Data-mining Pursues Identity and Differences. *Theory, Culture & Society*.
- Mackenzie, D. (2014). *A Sociology of Algorithms High-Frequency Trading and the Shaping of Markets*. Edinburgh: University of Edinburgh.
- Madsen, M. M., Blok, A., & Pedersen, M. A. (2018). Transversal collaboration: an ethnography in/of computational social science. In D. Nafus, *Ethnography for a Data-saturated World*. Manchester: Manchester University Press.
- Marchese, F. (2013). Tables and Early Information Visualization.
- Maurer, B. (2013). Transacting ontologies: Kockelman's sieves and a Bayesian. *HAU: Journal of Ethnographic Theory*.
- McCarthy, M. (2017). *Enacting the Semantic Web: Ontological Orderings, Negotiated Standards, and Human-machine Translations*. Milwaukee: University of Wisconsin-Milwaukee.
- McNally, R., & Mackenzie, A. (2012). Understanding the 'Intensive' in 'Data Intensive Research': DataFlows in Next Generation Sequencing and Environmental Networked Sensors. *The International Journal of Digital Curation*.

- MIT Technology Review. (2015, 2 23). *Computational Anthropology Reveals How the Most Important People in History Vary by Culture*. Retrieved from
<https://www.technologyreview.com/s/535356/computational-anthropology-reveals-how-the-most-important-people-in-history-vary-by/>
- Mitchell, T. (1997). *Machine Learning*. Redmond: McGraw-Hill.
- Mitchell, T. (2017). Key Ideas in Machine Learning.
- Miyazaki, S. (2012, 9 28). *ALGORHYTHMICS: UNDERSTANDING MICRO-TEMPORALITY IN COMPUTATIONAL CULTURES*. Retrieved from Computational Culture: A Journal of Software Studies: <http://computationalculture.net/algorhythmics-understanding-micro-temporality-in-computational-cultures/>
- Morita, A. (2014). The Ethnographic Machine: Experimenting with Context and Comparison in Strathernian Ethnography. *Science, Technology, & Human Values*, 214-235.
- Moss, M. (2017, 9 26). *Crowdwork for Machine Learning: An Autoethnography*. Retrieved from
<http://blog.fastforwardlabs.com/2017/09/26/crowdwork-for-ml.html>
- Murphy, K. (2012). *Machine Learning: A Probabilist Perpseptive*. Cambridge: The MIT Press.
- Mutzel, S. (2015). Facing Big Data: Making sociology relevant. *Big Data & Society*, 1-4.
- Nafus, D. (2018). *Ethnography for a Data-saturated World*. Manchester: Manchester Univeristy Press.
- Nafus, D. (2018). Working Ethnographically with Sensor Data. In D. Nafus, *Ethnography for a Data Saturated World*. Manchester: Manchester University Press.

Nafus, D., & Rattenbury, T. (2018, 3 7). *Data Science and Ethnography: What's Our Common Ground, and Why Does It Matter?* Retrieved from <https://www.epicpeople.org/data-science-and-ethnography/>

Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Norvaisas, J., & Karpfen, J. (2014). Little Data, Big Data and Design at LinkedIn. *EPIC*.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing.

Patel, N. (2013). For a Ruthless Criticism of Everything Existing: Rebellion Against the Quantitative-Qualitative Divide. *EPIC*, 43-60.

Physics arXiv. (2014, 8 14). *When A Machine Learning Algorithm Studied Fine Art Paintings, It Saw Things Art Historians Had Never Noticed*. Retrieved from <https://medium.com/the-physics-arxiv-blog/when-a-machine-learning-algorithm-studied-fine-art-paintings-it-saw-things-art-historians-had-never-b8e4e7bf7d3e>

Plessis, E., & Lorway, R. (2016). What really works: Understanding the role of “local knowledges” in the monitoring and evaluation of a maternal, newborn and child health project in Kenya. Monitoring and Evaluation. In S. Bell, & P. Aggleton, *Health and Social Development: Interpretive and Ethnographic Perspectives* (pp. 47-62). New York: Routledge.

Powell, A. (2017). Data Walks and the Production of Radical Bottom-up Data Knowledge. *International Communications Association*.

Rattenbury, T., Nafus, D., & Anderson, K. (2008). Plastic: a metaphor for integrated technologies. *UbiComp*.

Ribes, D., & Bowker, G. (2009). Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization*.

Rieder, B. (2017). Scrutinizing an algorithmic technique: the Bayes classifier as interested reading of reality. *Information, Communication & Society*, 100-117.

Roark, K. (2018). Participatory Big Data Ethics: Against AI Gaydar and Other Creepy Machines. *Society for Applied Anthropology*.

Seaver, N. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*.

Seaver, N. (2015). Bastard Algebra. In T. Boellstorff, & B. Maurer, *Data, Now Bigger and Better* (pp. 27-46). Chicago: Prickly Paradigm Press.

Seaver, N. (2015). The nice thing about context is that everyone has it. *Media, Culture & Society*.

Selman, B. (2014, 10 30). *Why Do We Conduct Qualitative User Research?* Retrieved from Mozilla UX: <https://blog.mozilla.org/ux/2014/10/why-do-we-conduct-qualitative-user-research/>

Seyfert, R., & Roberge, J. (2016). *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*. London: Routledge.

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*.

Simonite, T. (2018, 2 8). *Should Data Scientists Adhere to a Hippocratic Oath?* Retrieved from

<https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/#ampshare=https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath>

Sinders, C. (2017, 3 23). *Caroline Sinders on Ethical Product Design for Machine Learning.*

Retrieved from Design.Blog: <https://design.blog/2017/03/23/caroline-sinders-on-ethical-product-design-for-machine-learning/>

Sinders, C. (2017, 7 24). *The Most Crucial Design Job Of The Future: What is a data*

ethnographer, and why is it poised to become so important? Retrieved from

<https://www.fastcodesign.com/90134155/the-most-crucial-design-job-of-the-future>

Slobin, A., & Cherkasky, T. (2010). Ethnography in the Age of Analytics. *EPIC*.

Solanki, A., & Tewari, S. (2016). #GoingEthno in the Indian Bureaucracy. *EPIC*.

Solomoff, R. (1956). An Inductive Inference Machine.

Suchman, L. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions.*

Cambridge: Cambridge University Press.

Suchman, L. (2014). Human-Machine Autonomies.

Taylor, A. (2009). Machine Intelligence.

Thacker, E. (2005). *The Global Genome*. Cambridge: The MIT Press.

The social life of Learning Analytics: cluster analysis and the performance of algorithmic

education. (2016). *Learning, Media and Technology*, 3-16.

Thomas, M., & VELDHUIS, D. (2017, 10 11). *Learning to Trust Machines That Learn*.

Retrieved from <https://www.sapiens.org/column/machinations/game-theory-anthropology/>

Thomas, S., Nafus, D., & Sherman, J. (2018). Algorithms as fetish: Faith and possibility in algorithmic work. *Big Data & Society*, 1-11.

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*.

Vertesi, J. (2016, 7 7). *What robots in space teach us about teamwork: A deep dive into NASA*.

Retrieved from <http://ethnographymatters.net/blog/category/categories/co-designing-with-machines/>

Wallach, H. (2016, 7 15). *Machine Learning for Social Science*. Retrieved from SciPy 2016:
<https://www.youtube.com/watch?v=oqfKz-PP9FU>

Wang, T. (2013, 5 13). *Why Big Data Needs Thick Data*. Retrieved from
<https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7>

Wang, Y., & Kosinski, M. (2017). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *PsyArXiv*.

Wheeler, S. (2018, 2 14). *Data is a stakeholder*. Retrieved from
<https://towardsdatascience.com/data-is-a-stakeholder-31bfdb650af0>

Wilf, E. (2013). Toward an Anthropology of Computer-Mediated, Algorithmic Forms of Sociality. *Current Anthropology*, 716-739.

Ziewitz, M. (2013). What does transparency conceal? *Privacy Research Group*.

Ziewitz, M. (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values*, 3-16.

Ziewitz, M. (2017). A not quite random walk: Experimenting with the ethnomethods of the algorithm. *Big Data & Society*.