Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December 2019: 1–12

DOI: 10.1177/2053951719863819
journals.sagepub.com/home/bds
SAGE

Algorithms as folding: Reframing the analytical focus

Francis Lee[1], Jess Bier[2], Jeffrey Christensen[3], Lukas Engelmann[4], Claes-Fredrik Helgesson[1] and Robin Williams[4]

Abstract

This article proposes an analytical approach to algorithms that stresses operations of folding. The aim of this approach is to broaden the common analytical focus on algorithms as biased and opaque black boxes, and to instead highlight the many relations that algorithms are interwoven with. Our proposed approach thus highlights how algorithms fold heterogeneous things: data, methods and objects with multiple ethical and political effects. We exemplify the utility of our approach by proposing three specific operations of folding—*proximation*, *universalisation* and *normalisation*. The article develops these three operations through four empirical vignettes, drawn from different settings that deal with algorithms in relation to AIDS, Zika and stock markets. In proposing this analytical approach, we wish to highlight the many different attachments and relations that algorithms enfold. The approach thus aims to produce accounts that highlight how algorithms dynamically combine and reconfigure different social and material heterogeneities as well as the ethical, normative and political consequences of these reconfigurations.

Keywords

Actor-network theory, algorithms, folding, normativities, social theory, theory

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

Introduction

Algorithms appear able to connect different data, methods and objects smoothly between different settings, from matters of social distinction to natural catastrophe and crime.[1] The widespread introduction of algorithms in society seems closely tied to this ability to connect things that were previously unrelated. The attraction of algorithms thus often hinges on their ability to bridge the particularities of one setting to reshape and perform things in new manners (Ruppert, 2013a). Yet, the connective and bridging capacity of algorithms is little analysed. Rather many analysts today tend to frame issues of power and injustice in terms of bias within algorithmic systems (cf. Eubanks, 2018; Noble, 2018; Steiner, 2012).

In this article, we therefore propose to pay attention to algorithmic processes of connecting, relating or *folding*. The purpose of this is twofold: First, we propose a mode of analysing algorithms which directs attention to *operations of folding* over assessing the biases and opacity of algorithms. Second, we demonstrate the usefulness of this approach in understanding how society and nature are ordered *with* algorithms rather than *by* algorithms. That is, algorithms are in society, they do not control society.

Importantly, in proposing this mode of analysis, we attempt to move away from a focus on the hidden biases in algorithms or data (Angwin et al. 2016; Introna and

[1]Uppsala University, Uppsala, Sweden
[2]Erasmus University, Rotterdam, Netherlands
[3]Linkopings Universitet, Linkoping, Sweden
[4]University of Edinburgh, Edinburgh, UK

Corresponding author:
Francis Lee, Uppsala University, Uppsala, Sweden.
Email: francis@francislee.org

Wood, 2004; Kirkpatrick, 2016; Sandvig et al., 2016) as well as from a problematisation of algorithms in terms of opacity or accountability (Burrell, 2016; Diakopoulos, 2016; Zarsky, 2015).[2] We instead wish to highlight how algorithms relate and order a multitude of things—for example, different types of data, materials, methods, times, places or social relations—with sometimes unpredictable consequences.[3]

To be more concrete, it has become commonplace in the literature on algorithms to argue that algorithms could sustain, automate and accelerate oppression (Noble, 2018) or injustice (Eubanks, 2018) as well as reproduce social norms and bias (Steiner, 2012). A well-known case has been the introduction of algorithmic templates into sentencing and parole over the last decades in the USA. The hope was that the growth of databases of crime patterns and the statistical evaluation of re-offending rates would lead to evidence-based sentencing. In this way, the introduction of algorithmic sentencing was supposed to avoid the risk of biases associated with individual judgements in traditional judicial processes. However, in 2016, ProPublica, a journalistic NGO, evaluated the risk scores generated by one such algorithmic system, widely used within the US criminal justice system (Angwin et al., 2016). The evaluation showed that the risk scores tended to violate formal non-discrimination legislation as the system perpetuated the social and racial stratification of the incidence of crime and of convictions (Kirkpatrick, 2016).

One line of reasoning in this critical research implies that if only algorithms were designed in the optimal and correct way, they would generate results that were objective and fair. It is precisely this rule-bound and routinised nature of algorithms that seems to promise unbiased and fair sentencing. We find this reasoning misleading as it hides the multitude of relations algorithms are part of and produce. In a sense, the very notion of biased algorithms is linked to an objectivist understanding of how knowledge is produced, and worryingly sidesteps decades of research on the practices of knowledge production.[4] In this article, we instead want to stress that algorithms cannot offer uniquely objective, fair and logical alternatives to the social structures of our worlds.

Instead, we argue that algorithms must be understood as sociotechnical systems (cf. Seaver, 2018). They link society, technology and nature in a mesh of relations. And it is through multiple operations of folding—of relating things—that they work: It is in the many practices of relating, constructing, tinkering and applying that algorithms gain their power to reshape things. But, importantly in this perspective, it is not always the algorithm that is doing the shaping or folding. Sometimes humans fold things into the algorithm, and sometimes algorithms fold things into something else. Hence, agency is not fixed with the algorithms or with the humans (cf. Callon and Law, 1995). Thus, we argue that paying attention to processes and operations of folding can be a key mode for researchers to grasp and account 'for the distribution and fragmentation' of agency in algorithmic practices (Ruppert, 2013b: 272).

Consequently, we suggest that an analytical approach focusing on folding—on relating things that were previously unconnected—is better able to account for the varied processes by which algorithms order society and nature.[5] We consider case studies of the social and cultural impact of specific, and sometimes biased, algorithms as important inroads to understanding their effects, but we also want to stress the urgency of producing conceptual tools that can be used to analyse what algorithms do across multiple local and specific applications.[6] Folding thus provides a means of addressing efforts to 'dispel the algorithmic sublime' (Ames, 2018) in algorithmic studies. With this we want to contribute to going from 'myth to mess,' as Ziewitz puts it, and allow for an engagement with the myriad of ways that algorithms both order and reorder the world (2015: 6).

Analysing algorithms as an operation of folding

As we have stated above, we believe that a focus on operations of folding is a fruitful way of sidestepping both debates about the fairness and the opacity of algorithms. Thus, instead of mobilising the sometimes misapprehended metaphor of the 'black box' to uncover hidden and opaque operations of power within inaccessible algorithms (Pasquale, 2015), we are interested in the ways in which algorithms are part of ordering the social, natural and normative (cf. Mol and Law, 1994).

In wielding the fold as an analytical tool, we take inspiration from Bruno Latour's wide-ranging and diverse work on rhizomatic and relational ontologies, expressed through concepts such as folding, translation, rhizomes or networks (1999, 2002). Importantly for us, Latour has developed the notion of folding as a critique of essentialism that allows us to inquire into the mundane power of facts and artefacts. Drawing on Deleuze and Tarde, Latour has integrated the fold into his description of an ontology based in monadology. While for Deleuze the fold has become an important aspect of his work on difference and multiplicity, Latour uses the fold to describe associations and substitutions made by human and nonhuman actors that constitute the networks they operate within (Deleuze, 1993; Latour, 2010; Latour et al., 2012).[7]

To mobilise a useful figure, we draw on Michel Serres' and Bruno Latour's (1995) dialogue about a crumpled handkerchief to think about folding. In their conversation, they develop the folds of the crumpled handkerchief into a critique of a traditional linear view of time. Extending this metaphor, we might think about relations as becoming folded or torn, like the handkerchief, to encourage thinking in alternative topologies (cf. Mol and Law, 1994). Rather than thinking about objects, relations and concepts as stable entities with fixed distances and properties, we might attend to how different topologies produce different nearnesses and rifts. In this way, technologies, such as algorithms, can be understood as folding time and space as much as social, political and economic relations (cf. Latour, 2002: 248–249). By analysing algorithms in this manner, we argue that we can gain a better understanding of how they become part of ordering the world: sometimes superimposing things that might seem distant and sometimes tearing apart things that might seem close.

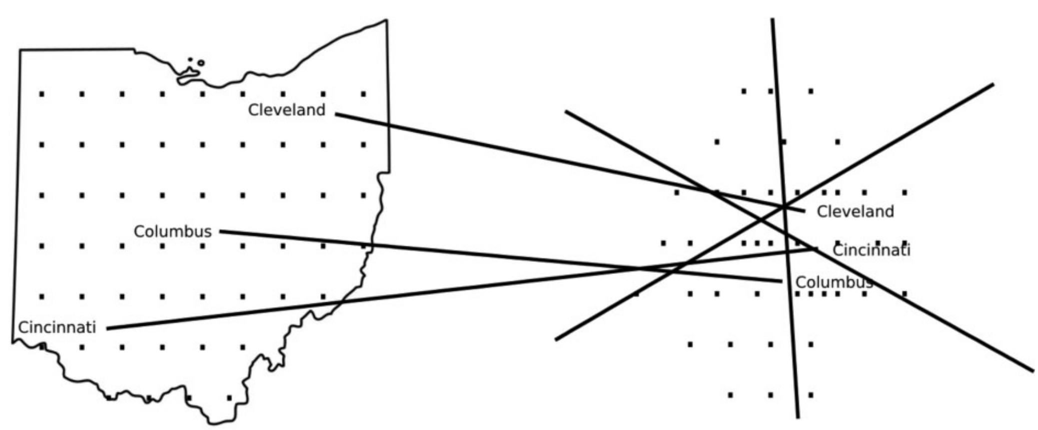
To be more concrete, using operations of folding to understand algorithms allows us to pay attention to how diverse things such as values, computations, datasets or analytical methodologies are algorithmically brought together to produce different versions of the social and natural orders. For example, a mathematical formula for aftershock prediction might be folded into a system for predictive policing (Benbouzid, 2017) or health statistics from the USA in the 1960s might be folded into German health recommendations in the 2010s (Bauer and Amelang, 2016).[8] Different times, places, computational strategies and versions of the social becoming folded together through the operations of algorithms.[9]

Analysing folding in four vignettes

To show the usefulness of paying attention to operations of folding, we analyse four empirical vignettes where algorithms help order society and nature. The vignettes reflect work done by the authors in diverse settings and go into different levels of empirical detail. In analysing these settings, we bring together some ways in which algorithms fold sets of data, modes of reasoning and objects and subjects. To set them apart from the general argument, the vignettes are placed in boxes, and interspersed with analytical commentaries that draw out our main argument. The vignettes are illustrative of some facets of the operations of folding, and each vignette highlights a particular theme. Importantly, these vignettes have been chosen to demonstrate how algorithmic operations of folding work in practice, from producing proximities and universals, to bringing these normative universals to bear on individuals.

Proximation: From proximities of social groups to proximities of transmission

Our first vignette deals with the history of mapping AIDS. The algorithmic generation of a novel ‘AIDS space,’ as designed by the geographer Peter Gould, draws attention to how algorithms can rearrange a geographic complexity into a non-geographical topography. Here we attend to how an algorithmic transformation of an AIDS visualisation can shift epidemiological attention from the *populations* that were deemed most at risk toward the *regions* that are most likely to be affected. It did so by replacing one normative framework of proximity with another. The picture of an epidemic tidal wave sweeping over the country was replaced with a map that instead reflected the spatial coordinates of behaviour and identities characteristic for AIDS. The traditional view was that homosexual men, heroin users, haemophiliacs and Haitians were the origin of the epidemic, but Gould’s AIDS space instead crafted a spatial representation which highlighted the specific patterns in which the epidemic worked, producing new proximities and distances to the AIDS epidemic.



Drawn after diagram in Gould et al. (1991: 86).

Vignette 1: Folding different versions of proximity/mapping the AIDS space

In 1990, the geographer Gould expressed his discontent with how the AIDS epidemic was mapped across the United States (Gould et al., 1991). Gould was not satisfied with the image of AIDS conjured in the traditional sequence of maps. These maps usually showed the progress of the disease over time like a tidal wave washing over national geographies in a sequence of steps. These geographical and temporal visualisations, he argued, allowed for complacency regarding the spatial pattern he had observed, which was not comparable to a slow and homogenous spreading.

To solve his problem, Gould developed a competing algorithm, which would capture the pathways and complicated spatial–temporal distribution of AIDS. He translated the geographic distribution and the inhabited landscape into a statistical representation of the rapid transmission of the emerging epidemic. As a result, the map was replaced by a diagram in which the spatial distribution had become a characteristic of the epidemic—as the epidemic now became visualised as a cluster centred around urban habitation.

Gould designed his algorithm as a way to predict the next outbreak. He was convinced that sequential series of maps could only deliver a vague picture of threat which might shore up a sense of false security. Moving away from the evocative image of a tidal wave, Gould's team aimed to integrate the highly specific social structure of the epidemic and its relationships to urban nodal points. His intent was to alarm teenagers, students and health practitioners who did not acknowledge their own proximity to the epidemic. A new set of proximities were forged through Gould's topography.

In contrast to the traditional tidal wave, Gould's new algorithm laid out a model for rethinking the distribution of AIDS with respect to relative population density. The argument was that AIDS could be differentiated from a contagious disease like the plague, for which diffusion follows a gradual distribution over geographical space, reaching village after village as if it were a map of an extending flood. Gould's maps instead showed how AIDS jumped from one large city to another, accompanied by slower diffusion to the surrounding countryside. This crafted a geographic projection, in which the disease was not plotted in relation to the space in which it moves, but rather space was rearranged along the characteristic dynamics of the epidemic. Gould plotted what he called an 'AIDS space.' By moving the urban centres out of their geographic position and grouping them together according to the probability of the next infection, he could visualise the proximity of the next AIDS event (Engelmann, 2018: 124; Gould et al., 1991; Koch, 2005: 272).

The 'AIDS space' provides insight into how algorithms can fold the world to create new proximities. Gould's algorithm produced a new order of the epidemic built on its transmission patterns and associated risk behaviours, and plotted a map of AIDS as a new topological order, which was designed to enable an accurate prediction of the epidemic. The previous focus on the *proximity of particular populations* to the epidemic was thus replaced with a focus on the specific *patterns of transmission* and risk. Gould thus dissolved the geographical distance of the cities affected by AIDS. He used his algorithm to draw a map entirely different from the usual visual representations: his map transformed the geography of the USA into *a new spatial distribution* that was deemed more characteristic of AIDS.

Gould's algorithm takes on a double function in this context. First, the algorithm re-assembles the transmission pathways characteristic for HIV and presents a formalised expression of the nature of AIDS. Its first impact was to replace a focus on particular risk groups with a focus grounded in the formalisation of the epidemic as a series of infections. Gould's algorithm thus transformed sexual behaviours and practices into a new set of proximities. But second, the algorithm took these characteristic patterns of the epidemic and re-shaped them into a new spatial pattern, transforming its social topography of infection into a geography of transmission in which *new proximities and new spaces of risk were made visible*.

Gould's AIDS space became a timely reminder that social and cultural framings of the epidemic had constrained the understandings of both the research community and the general public. It was intended to replace the traditional picture, which was attached to stereotypes, rumours and false epidemiological assumptions. Thinking AIDS through its unique spatial pattern was an invitation to unsee the proximity of homosexual men, heroin users, haemophiliacs and Haitians to the epidemic. Instead, Gould's map evoked a picture of a new spatial order—a set of *social* proximities was replaced with a set of *spatial* proximities. Two versions of the AIDS epidemic were set against each other.

This brings us to our second operation of folding: the algorithmic production of universals through a heterogeneity of particulars. Here we attend to the folding of a global universal, from a multitude of elements, through an algorithm that was used to produce the 'Current Zika State.'

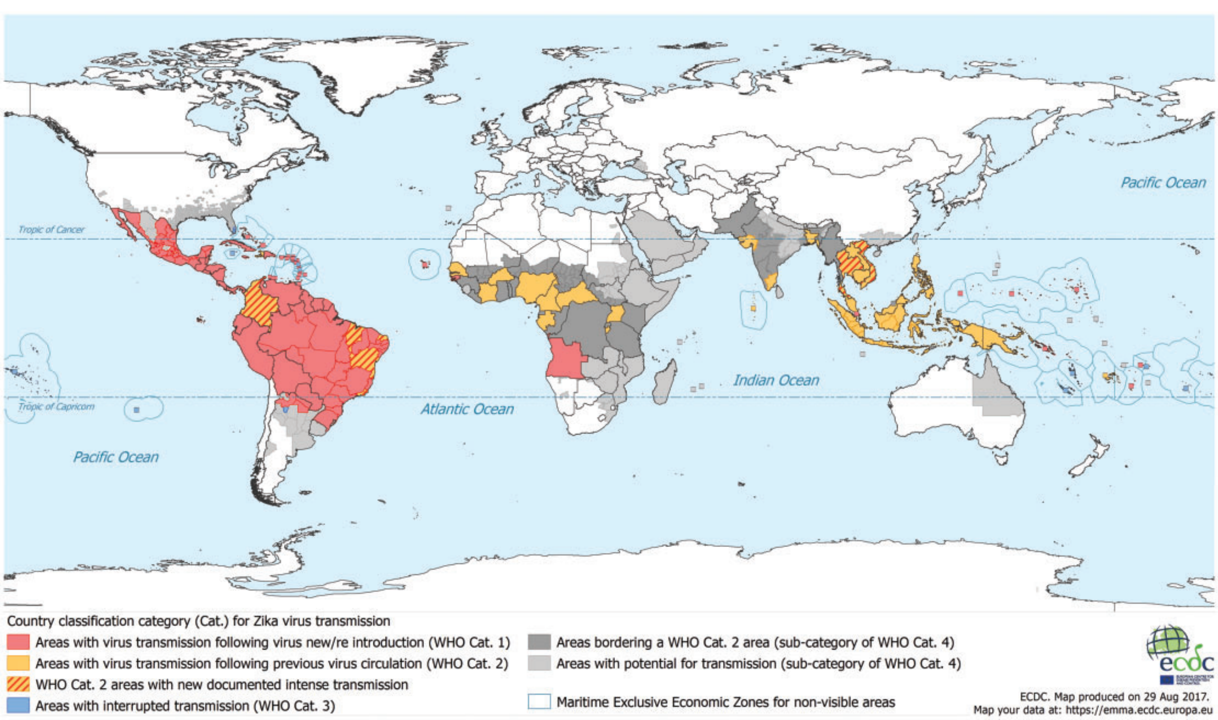
Universalisation: From a multitude of particulars to a global neverwhere

The algorithmic production of the 'Current Zika State' shows how algorithms transform a set of local particularities into an apparent global universal, which also performs certain places as proximate to the Zika epidemic. It demonstrates—through the construction of the 'Current Zika State'—how a series of particular data, measurements, calculations and hypotheses are algorithmically assembled and merged to project a universal view of Zika. These operations give particular times and places the ability to stand for all times and places. Far from existing outside of or exterior to particularity, universals like those of the global Zika map, or the AIDS space described above, complexly combine, incorporate and interiorise particular data and calculations from different times and places. Here we deal with an operation of folding where a heterogeneous set of partial elements is brought together and transformed to produce a universal view. A new universality is created that appears to be self-evident—a natural fact of the world.

Vignette 2: A global neverwhere/producing the Current Zika State

The goal of disease surveillance is to control the spread of disease. Algorithms, machine learning and databases promise to handle larger and larger sets of data—and more data promises more sensitivity to disease outbreaks. Zika is a recent addition to the global bestiary of pandemic threats, and quickly rose to fame before the Olympic games in Rio de Janeiro. Zika provoked a flurry of media attention. Media headlines such as 'Zika Virus Makes Rio Olympics a Threat in Brazil and Abroad' circulated the globe (Kassam, 2016). The fact that Zika is both sexually transmitted and transmitted by the *Aedes aegypti* mosquito triggered a scare that the disease would spread rapidly across continents.

The aim of government disease surveillance organisations is to track, prevent, and curtail different epidemics in the world, including Zika. Surveilling any disease depends on a huge amount of work, and Zika is no different. Zika surveillance depends both on quantifying Zika cases around the world as well



The Current Zika State.[10]

as other infrastructures for quantified risk prediction. One of the challenges is that many warmer and wetter countries do not have the resources to build and maintain infrastructures for tracking Zika and the feared *Aedes aegypti* mosquito. How do you then capture where there is Zika risk globally?

To address these challenges, the European Centre for Disease Control and Prevention (ECDC) created an algorithm to track the Zika epidemic and to predict Zika risk across the globe. This algorithm drew together several different datasets, computational methodologies and infrastructures that tied the Zika epidemic to both the modelling of mosquitos and climate zones. For example, the ECDC algorithm utilised a risk modelling approach to predict the presence of the *Aedes aegypti* mosquito in a geographical area. This risk model harnessed data about where the *Aedes aegypti* had been found, taken from different infrastructures, times and places across the globe. For instance, the geographical range of the *Aedes aegypti* was calculated based on data from the US Centers for Disease Control and Prevention, but was also based on data about the mosquito published in scientific journals.[11] This risk model of Zika also folded historical climate data drawn from a multitude of weather satellites, and computations from several different climate models. In sum, the risk model aimed to predict whether a habitat could be suitable for the *Aedes aegypti* mosquito by combining data from many different times and places.

However, the ECDC algorithm did not solely tie the Zika risk to computations pertaining to the *A. aegypti* mosquito. Zika risk was also inferred by modelling the risk of Dengue (which is also transmitted by the *A. aegypti* mosquito) as well as by using a so-called Köppen–Geiger climate classification of the world. All of these different models, datasets, classifications and computations were harnessed in the Zika algorithm to produce a snapshot of what was published as the 'Current Zika State' of the world. The point is that the 'Current Zika State' drew on a plethora of different times, places, computational efforts and infrastructures. To know the risk of Zika, algorithms connected past, present and future as well as a multitude of particularities of Zika.

The ECDC algorithm, just as the AIDS maps above, brings together a multitude of datasets to produce a universal and global view of a pandemic, where certain countries and people are more proximate to the Zika epidemic than others. However, what seems to be a map of disease encompassing the entire world is actually a chimera made up of very different data. This global and universalising map ignores absences in the data and the mosaicked qualities that come from the multitude of different data forms. The particularity and partiality of the data are removed from the global view. From the map itself, it is not clear how and why different data are combined, and for which areas. The map gives an account of Zika transmission, but it contains no traces of the work that was necessary to collect these data and combine them into a global view of Zika.

These heterogeneities represent a diversity of practices, locations and timescales, bringing a multitude into a universal coherence. In short, the Zika algorithm is an excellent example of *universalisation*: Through the algorithm's different operations of folding things together, particulars are transformed into apparent universals. This is the apparent Janus-face of the algorithm: a complex and heterogeneous past, which can produce a smooth and universal present or future. A set of particularities becoming a smooth and coherent 'view from nowhere' (cf. Haraway, 1988)—an algorithmic neverwhere.

Normalisation: From enveloping to developing the normal

We now turn to the algorithmic production of 'the normal' by attending to the prediction of stock market risk. In finance, just as in many complex systems, regular activity is characterised by its unpredictability. It can therefore be exceedingly difficult to determine precisely what the normal state of the financial system is. Perhaps because of this unpredictability, algorithms have incredible justificatory power in debates over whether a particular economic pattern represents normal or abnormal variation of economic activity. There are currently numerous efforts to algorithmically detect aberrant patterns that diverge from ordinary background noise of 'normal' economic activity.

In recent years, debates over the normal state of financial markets have focused on how aberrations arise. Algorithmic models are routinely used to argue that financial crashes are normal to markets, on the one hand, and that they are abnormal and aberrant, on the other. Economists on each side put forth different algorithms and prioritise different styles of reasoning, from statistical judgement to the recognition of visual patterns. Different algorithms are thus built to identify deviations and abnormalities based on particular versions of normality. These versions of normality are expressed through mathematical functions such as, for instance, the normal or 'Bell' curve. Thus, in algorithms built to detect normality and abnormality, specific versions of the normal are translated into mathematical form and folded into the calculative logic of particular algorithms.

Vignette 3: Competing versions of the normal/ modelling stock market risk

Financial markets do not collapse every day, but they do collapse, and their crashes are unpredictable. Attempts to foresee crashes through algorithms depend in large part on different conceptions of what markets are and how they work. Today, two ways of understanding markets are common. First, there is the dominant view, which sees crashes as outliers: rare and unlucky events. Second, there are alternative perspectives, which see crashes as integral to contemporary capitalist markets: a likely, if unpredictable, occurrence.

The Black–Scholes–Merton model (BSM) is one of the most well-known examples of the dominant paradigm that sees markets as outliers (MacKenzie, 2006). Like many mainstream models, it relies on the normal or 'Bell' curve, which implies that small changes in markets are incredibly common and very large changes, i.e. crashes, are incredibly rare. Thus, the BSM model includes as one of its assumptions that major crashes are unlikely in contemporary financial systems. In contrast, alternative models like those of Benoit Mandelbrot (Mandelbrot, 1997; Mandelbrot and Hudson, 2004) avoid the normal curve, instead relying on graphs like the power law curve.

Unlike the normal curve, the power law curve includes the assumption that very large changes in the market occur far more often than traditional models, based on the normal curve, would suggest. As a result, models that use the power law curve include the assumption that crashes are in fact normal, in the sense of frequent, occurrences.

Both the dominant BSM models and Mandelbrot's alternative models rely on algorithms, but the two types of model involve fundamentally different assumptions about what is normal. So the choice of which model to use necessarily involves a choice about which kind of normal—whether crashes are rare or common—one should assume.

Yet, algorithms cannot tell us which choice to make because the decision about what is normal is central to deciding which algorithm to use in the first place. Instead, the choice over assumptions about the normal is made using a variety of styles of reasoning including statistical knowledge, previous use of algorithms, professional familiarity with trading practices, systemic knowledge of financial regulation, discussions with peers, and so on. So contrary to the presumption that the use of algorithms would resolve what is normal, the algorithms make different conceptualisations of the normal even more complex.

Mandelbrot's model was intended, in part, to settle debates over what is normal for financial markets.

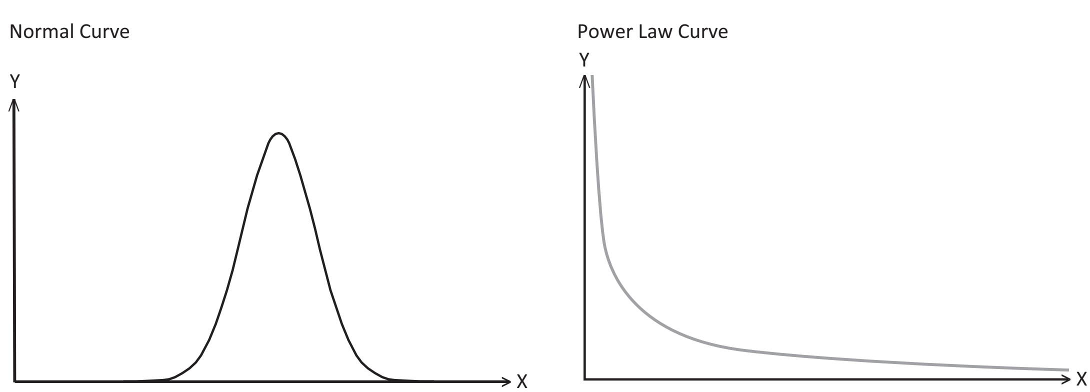


Figure: The normal curve and the power law curve. The x-axis indicates the magnitude of a particular change—for example, the extent of the rise or fall of a stock market in a given time period. The y-axis indicates the frequency of changes of that magnitude—for example, how often the market has risen or fallen that amount. Given the same set of parameters, a process modeled with the normal curve approaches the x-axis more quickly than the power law curve, indicating that there are fewer changes that are either extremely large or extremely small.[12]

However, to date, far from resolving conflicts over the normal, Mandelbrot's model simply adds another definition of the normal into the mix.

Thus, for stock market models, there is competition between the different normals and different algorithms. Practical assessments of markets, including the kinds of judgements about 'how well' the market is doing, rely on both BSM and Mandelbrot algorithms, which are each enfolded with different ideas about a normal market. The use of algorithms in finance thus involves transforming different versions of the normal, including statistical norms, social knowledge about the frequency of market crashes and visual assessments of the normal appearance of a graph, into the overall production of what is normal for finance.

This is not unique for stock market models. Most algorithms are folded with particular versions of the normal. For instance, ideas about normality were also folded into the Zika algorithm. While modelling the habitat of the *Aedes aegypti* mosquito, environmental data stemming from satellites were mathematically transformed into oscillating cyclical curves. The assumption built into the mosquito-model was that the normal behaviour of nature was cyclical in terms of, for instance, rainfall, temperature or vegetation index. However, this cyclical version of 'normal nature' does not fit well with non-cyclical changes in the environment—such as climate change, deforestation or processes of urbanisation. Likewise, in the case of Gould's AIDS space, what was traditionally represented as an epidemic tidal wave was replaced by a new mathematical description of the normal transmission patterns of the AIDS epidemic. Different versions of normality were folded with the different algorithms.

This indicates that algorithms alone cannot settle debates about the state of the world. Rather than being the source of well-defined normalities, algorithms are constantly folded with different valuations and styles of reasoning in producing what is considered normal. Consequently, algorithms are used in struggles over what is normal, and are often used in ways that complicate, rather than resolve, debates over normality.

Bringing it all together: Proximations, universalisations and normalisations in the Recent Infection Testing Algorithm

This brings us to our last vignette, where we bring together our three operations of folding—*proximations*, *universalisations* and *normalisations*—in one setting. Here, we turn to a second algorithm related to AIDS, a Recent Infection Testing Algorithm (RITA).[13] RITA was first developed to estimate the incidence rates of HIV by calculating the 'recency' or major time-points of an infection process in a population. But to complicate the matter, RITA is also sometimes used to do epidemiological assessments for individual patient management. Thus, this algorithm is, among other things, used for bringing the aggregated population dynamics of the AIDS epidemic to bear on an individual patient's disease.

However, the estimated time-points of infection have limited levels of reliability and robustness, and their applicability for individual cases is unclear, as the calculated time-points are merely a statistical approximation. The estimated time-points can be said to be a one-size-fits-all approximation of what a normal immune response to HIV is, based on a particular statistical population. To compound the issue, RITA does not incorporate individual case details, nor the myriad of potential exceptions to the existing norm, into the approximations. RITA thus assembles a system in which a statistical pattern produces a statistical view of the progression of a 'normal AIDS infection.' These computed statistical time-points are, as we show in the vignette, then sometimes brought to bear on individual patient assessments and plans for future treatment. A universalised population, and an algorithmic enactment of the progression of a normal AIDS infection, is thus brought proximate to individual patients.

Vignette 4: How a population algorithm became an algorithm for assessing individuals

Algorithmic practices have become entangled with AIDS and HIV in a variety of ways. In the domain of HIV governance, an algorithm can both be understood as an entity that coordinates a testing process—often through visualisation and images—as well as an object that calculates and formats the results of different laboratory tests. RITA is an example of such a device. RITA was first designed for use in public health practices, specifically to calculate the incidence of HIV infections. The goal was to calculate the recency of infections in a tested population by statistically estimating the significant time-points or steps in the infection.[14]

However, since its origin as a device for population measurement, RITA has also become a tool for estimating how recently an individual was infected with HIV. As a consequence of this shift from the population to the individual, RITA may, for instance, be used to verify the timing of infection that a patient accounts for. Furthermore, as it is a punishable offence in certain jurisdictions to not inform a sex partner of being HIV infected, RITA can also be used to validate the testimony of a patient when prosecuting transmission cases. In such cases, the algorithmically computed progression of a 'normal infection' is folded onto an individual case, with potentially grave consequences for the individual.

In essence, the effect of RITA is that it transforms the temporality of HIV infection by staging some infections as 'recent' and others as 'long standing.' But recency is a complicated matter. The thresholds used to mark a recent infection may be statistically reliable on the population level, but might be diagnostically problematic on the individual level (Kassanjee et al., 2012). Immune system variation in patients, and other factors that are yet unknown, quite often produce RITA results that can be argued to be false using other techniques. This problem can be addressed through confidence intervals and modelling on the population level, but have far more severe implications when individual patients become accountable to them—such as when patients are prosecuted for transmitting the disease.[15]

While RITA still plays an important role in the national and international surveillance of AIDS populations, it has thus also become used for prevention planning, identification of individuals for research and managing individual patients living with HIV (Murphy et al. 2017). In these ways, the statistically produced AIDS population is folded onto individual AIDS patients.

Yet, the population constructed with RITA includes assumptions that do not apply to all patients, and this creates problems when applying RITA to the individual level. For example, one might think low levels of HIV would indicate a recent infection. Contrary to this logic, it has been shown that some patients, who have been identified as infected, suppress HIV to nearly undetectable levels—without medical treatment. They have been labelled 'elite controllers' by practitioners in the field and do not fit into the progression of a 'normal infection.' These elite controllers demonstrate that the assumptions about what is a normal infection across all AIDS populations cannot be taken for granted. Different individual infections can progress according to individual rates that do not correspond to the statistical estimates. So applying RITA's 'normal rate of progression' to an individual might actually mislead doctors.

The use of the RITA thus underscores the complex operations of folding through which algorithms can shape knowledge about and action on the world. Indeed, the RITA—just as the algorithms in our other vignettes—produces both universalisations and normalisations. It produces both a *universalised* AIDS population based on a plethora of data as well as a computed 'normal AIDS infection.' Hence, just as a financial algorithm produces a particular version of a normal market, RITA produces a particular version of a 'normal AIDS infection.' But the RITA also brings this 'normal AIDS infection' *proximate* to individual AIDS patients in that an individual's disease progression can be measured against the normal infection. Thus, algorithms can become a point where 'everything is tied together in one particular spot' (Serres and Latour, 1995: 87)—particulars become universals, universals produce normals, and new proximities are made.

A new direction in algorithm studies? Thinking with operations of folding

As algorithms are increasingly used to bring together heterogeneous data, methods, objects and relations, they also help to produce new orderings of society and nature. We have argued that paying attention to *operations of folding* can be a key strategy for understanding how a diversity of objects are refashioned through algorithmic practices, and that this strategy might broaden and complement approaches that assess algorithms for fairness or bias or lament their opacity (*pace* Angwin et al., 2016; Kirkpatrick, 2016). That is, we argue that when 'unbiased data' and 'fair algorithms' become the focus, there is a risk that questions about situatedness, partiality, and the production of the 'normal' become invisible. But they remain crucial questions to pose, if we are to deal analytically with the increasing influence of algorithms in society.

In proposing this approach, we emphasise that folding is not an innocent operation, and that algorithms do not work through neutral operations that bring the world together in a detached manner. Rather any analysis of algorithms needs to acknowledge that they work through *attachment* and *relation*, not through detachment, biases or objectivity. Thus, drawing on Latour's (2002) work, we argue that folding entails a *translation not a transmission,* in the sense that an algorithm does not fold things unaltered. To be clear, political relations and attachments can certainly come in the form of nefarious and hidden bias or calculations in the algorithm, but there are many other attachments and forms of politics that we need to heed in our analyses.

In attending to operations of folding above, our first analytical move was to zero in on how *algorithms make proximate* different objects and relations. We showed how an epidemiological model, which visualised an epidemic as a tidal wave, was replaced with a new topology that brought geography into focus rather than specific risk groups. What people, objects or relations are then produced as proximate or far away by algorithms?

Our second move was to analyse *how universals are produced* through the folding of partialities. By attending to the algorithmic construction of a global disease map, we zeroed in on how a multitude of heterogeneous objects—data, methods, objects and relations—were used to assemble a global and coherent map of disease. What partialities are then made to stand in for the whole? What is made part of the universal and what becomes invisible?

Our third analytic entailed attending to how *algorithms are folded with different versions of the normal*. In this we ask: How do assumptions about the normal become folded into algorithms? And how is the normal or abnormal then performed with algorithms?

So where do we go from here? Analysing operations of folding means remaining open to the different types of relations, politics and attachments that are made and unmade with algorithms. It means tracing operations of folding, regardless of what is folded and by whom. It means remaining agnostic as to what things can be folded with algorithms, and in what ways they can be folded. It means following algorithms through a diverse array of practices, both social and technical, sometimes in the same place, but sometimes through different settings, different logics, and different times and places. Rather than reifying algorithms as uniquely powerful and opaque black boxes, analysing operations of folding opens a different route, which highlights how algorithms can dynamically combine and reconfigure different social and material heterogeneities. We can then begin in earnest to investigate the relations, ethics and politics of algorithms.

Acknowledgements

We would like to express our gratitude to all participants of the workshops of *The Algorithm Studies Network* between 2015 and 2018, in which this perspective has been developed, reviewed and refined. We further would like to thank our editors, Matt Zook and Evelyn Ruppert, as well as our anonymous reviewers for detailed and productive feedback.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We would like to thank *Riksbankens jubileumsfond* for their generous funding of the workshops and PhD summer school that made this article possible.

ORCID iD

Francis Lee https://orcid.org/0000-0002-7206-2046

Notes

1. To take two examples: algorithms are sometimes claimed to be able to 'automate social theory' by creating recommendation systems that do not rely on demographic data (Seaver, 2012); algorithms that were developed to detect earthquakes have also been re-purposed for use in systems for the so-called predictive policing—and are claimed to be able to predict where and when crime will happen (Benbouzid, 2017).
2. It is not possible in the limited space of a journal article to encompass the complete breadth and depth of research on algorithms, but a good starting point for the curious reader is: https://socialmediacollective.org/reading-lists/critical-algorithm-studies/
3. Importantly, we do not assume that operations of folding are exclusive to the domain of algorithms. One could, for example, argue that, historically, rules have often led to effects of folding, as they presume a reorganisation of social, cultural and natural orders. See, for example, Daston (2019).
4. See, for example, Bloor (1976) or Collins (1975) for classic examples. We argue that an analytical focus on operations of folding opens up a space for dealing with algorithmic effects as they are related to, or attached to, a heterogeneity of elements, without succumbing to the temptation of producing 'Whiggish histories' that uncover the unbiased reality that is hidden behind the algorithm.
5. Algorithms can do many different things. They can sort, filter, recombine. The list is almost endless. However, we argue that folding can capture a central and generic aspect of many of today's algorithmic systems.
6. This mode of analysis could for instance be used to deal with cases such as the one David Ribes describes when he examines the pursuit of an idealised 'domain-independence' in the historical development of data-science: a good program, a useful algorithm was thought to be capable of finding application across domains, such as medicine and law or education and biology (Ribes, 2018).
7. In approaching technology through the concept of the fold, Latour draws on Deleuze but pursues a much more pragmatic notion of the fold. Technology folds time, space and the type of actants involved. As such Latour defines a 'regime proper to technology by the notion of *fold*, without giving it all the Leibnizian connotations that Gilles Deleuze (1993) has elaborated so well.' (Latour, 2002: 248).
8. For more on the use of algorithms in predictive policing, see, e.g., Amoore (2013) and Ruppert (2013a).
9. Our proposed analysis of folding can thus be incorporated into existing research that interrogates algorithms in society. This includes practices through which algorithms are implemented (Christin, 2017), work on how

algorithms are reshaping observation (McQuillan, 2016) and rationality (Lowrie, 2017), while also addressing the relations between the algorithmic and the non-algorithmic (Dourish, 2016).

10. This version of the map was published at https://ecdc.europa.eu/en/publications-data/current-zika-transmission-worldwide on 29 August 2017.
11. To complicate the matter further, there are very few studies published about where the *Aedes aegypti* mosquito is *not* present. This lack of data is solved by simulating the *absence* of *Aedes aegypti* based on ecological distance.
12. As Donald MacKenzie has reminded us, in this figure the tails of the normal curve blend into the horizontal axis, though the Gaussian distribution's tails are asymptotic: they never actually get down to zero.
13. This algorithm is referred to as RITA by the World Health Organization, though this name is contested by the makers of the algorithm who prefer the label Test for Recent Infection.
14. While variations of these algorithms can be found, most versions of RITA include both immunological and virological components. These components quantify the strength of the immune system, the presence of viral genes in the sample population as well as a function that identifies subjects undergoing anti-retroviral treatment.
15. As noted by a recent paper to the *Global Commission on HIV and the Law* (Weait, 2011):

> In the case of HIV transmission, new tests, known as RITA (Recent Infection Testing Algorithm) tests, are sometimes being used to assess rates of recent infection in the population, and it is possible that a RITA test result for an individual sample might be offered as evidence of the timing of infection. These tests are sometimes also known as STARHS (Serological Testing Algorithm for Recent HIV Seroconversion) tests. Prosecutors should be aware that there are limitations on the reliability of such evidence at an individual level.

References

Ames MG (2018) Deconstructing the algorithmic sublime. *Big Data & Society* 5(1): 1–4.

Amoore L (2013) *The Politics of Possibility: Risk and Security Beyond Probability*. Durham, NC: Duke University Press.

Angwin J, Larson J, Mattu S, et al. (2016) Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica* 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Bauer S and Amelang K (2016) Calculating, reshuffling and translating population data: Epidemiological risk scores as accounting devices. In: *EASST/4S conference*, Barcelona, 31 August–3 September 3.

Benbouzid B (2017) On crime and earthquakes. Predictive policing at the crossroads between science, technology and divining. *Réseaux* 206(6): 95–123.

Bloor D (1976) The strong programme in the sociology of knowledge. In: Bloor D (ed.) *Knowledge and Social Imagery*. London: Routledge, pp. 1–19.

Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1): 1–12.

Callon M and Law J (1995) Agency and the hybrid Collectif. *South Atlantic Quarterly* 94(2): 481–507.

Christin A (2017) Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society* 4(2): 1–14.

Collins HM (1975) The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology* 9: 205–224.

Daston L (2019) *Against Nature*. Cambridge, MA: MIT Press.

Deleuze G (1993) *The Fold: Leibniz and the Baroque*. Minneapolis: University of Minnesota Press.

Diakopoulos N (2016) Accountability in algorithmic decision making. *Communications of the ACM* 59(2): 56–62.

Dourish P (2016) Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3(2): 1–11.

Engelmann L (2018) *Mapping AIDS. Visual Histories of an Enduring Epidemic*. Cambridge: Cambridge University Press.

Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Gould P, Kabel J, Gorr W, et al. (1991) AIDS: Predicting the next Map. *Interfaces* 21(3): 80–92.

Haraway D (1988) Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14: 575–599.

Introna LD and Wood D (2004) Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society* 2(2/3): 177–198.

Kassam A (2016) Zika virus makes Rio Olympics a threat in Brazil and abroad, health expert says. *The Guardian*, 12 May. Available at: www.theguardian.com/world/2016/may/12/rio-olympics-zika-amir-attaran-public-health-threat (accessed 3 July 2019).

Kassanjee R, McWalter TA, Bärnighausen T, et al. (2012) A new general biomarker-based incidence estimator. *Epidemiology* 23(5): 721.

Kirkpatrick K (2016) Battling algorithmic bias. *Communications of the ACM* 59(10): 16–17.

Koch T (2005) *Cartographies of Disease. Maps, Mapping, and Medicine*. Redlands: ESRI Press.

Latour B (1999) On recalling ANT. In: Law J and Hassard J (eds) *Actor Network Theory and After*. Oxford: Blackwell, pp.15–25.

Latour B (2002) Morality and technology: The end of the means. *Theory Culture & Society* 19(5/6): 247–260.

Latour B (2010) Tarde's idea of quantification. In: Candea M (eds) *The Social After Gabriel Tarde: Debates and Assessments*. Routledge, London, pp.145–162.

Latour B, Jensen B, Venturini T, et al. (2012) 'The whole is always smaller than its parts' – A digital test of Gabriel

Tardes' monads. *The British Journal of Sociology* 63(4): 590–615.

Lowrie I (2017) Algorithmic rationality: Epistemology and efficiency in the data sciences. *Big Data & Society* 4(1): 1–13.

MacKenzie D (2006) *An Engine, Not a Camera: How Financial Models Shape Markets*. Cambridge, MA: MIT Press.

Mandelbrot BB (1997) *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. New York: Springer.

Mandelbrot BB and Hudson RL (2004) *The (Mis)Behaviour of Markets: A Fractal View of Risk, Ruin and Reward*. London: Profile Books.

McQuillan D (2016) Algorithmic paranoia and the convivial alternative. *Big Data & Society* 3(2): 1–12.

Mol A and Law J (1994) Regions, networks and fluids: Anaemia and social topology. *Social Studies of Science* 24(4): 641–671.

Murphy G, Pilcher CD, Keating SM, et al. (2017) Moving towards a reliable HIV incidence test – Current status, resources available, future directions and challenges ahead. *Epidemiology & Infection* 145(5): 925–941.

Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Pasquale F (2015) *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.

Ribes D (2018) STS, meet data science, once again. *Science, Technology, & Human Values* 514–539.

Ruppert E (2013a) Not just another database: The transactions that enact young offenders. *Computational Culture* 3: 1–13.

Ruppert E (2013b) Rethinking empirical social sciences. *Dialogues in Human Geography* 3(3): 268–273.

Sandvig C, Hamilton K, Karahalios K, et al. (2016) When the algorithm itself is a racist. *International Journal of Communication* 10: 4972–4990.

Serres M and Latour B (1995) *Conversations on Science, Culture, and Time*. Ann Arbor: University of Michigan Press.

Seaver N (2012) Algorithmic recommendations and synaptic functions. *Limn* 1(2): Available at: https://limn.it/articles/algorithmic-recommendations-and-synaptic-functions/.

Seaver N (2018) What should an anthropology of algorithms do? *Cultural Anthropology* 33(3): 375–385.

Steiner C (2012) *Automate This: How Algorithms Came to Rule Our World*. New York: Portfolio.

Weait M (2011) The criminalisation of HIV exposure and transmission: A global review. In: *Working paper prepared for the third meeting of the technical advisory group, global commission on HIV and the law Third Meeting of the Technical Advisory Group, Global Commission on HIV and the Law*, 7–9 July 2011.

Zarsky T (2015) The trouble with algorithmic decisions. *Science, Technology & Human Values* 41(1): 118–132.

Ziewitz M (2015) Governing algorithms. *Science, Technology, & Human Values* 41(1): 3–16.

Big Data & Society
July–December 2019: 1–12
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951719858742
journals.sagepub.com/home/bds

Perspectives on algorithmic normativities: engineers, objects, activities

Jérémy Grosman[2] and Tyler Reigeluth[1]

Abstract

This contribution aims at proposing a framework for articulating different kinds of "normativities" that are and can be attributed to "algorithmic systems." The technical normativity manifests itself through the lineage of technical objects. The norm expresses a technical scheme's becoming as it mutates through, but also resists, inventions. The genealogy of neural networks shall provide a powerful illustration of this dynamic by engaging with their concrete functioning as well as their unsuspected potentialities. The socio-technical normativity accounts for the manners in which engineers, as actors folded into socio-technical networks, willingly or unwittingly, infuse technical objects with values materialized in the system. Surveillance systems' design will serve here to instantiate the ongoing mediation through which algorithmic systems are endowed with specific capacities. The behavioral normativity is the normative activity, in which both organic and mechanical behaviors are actively participating, undoing the identification of machines with "norm following," and organisms with "norminstitution". This proposition productively accounts for the singularity of machine learning algorithms, explored here through the case of recommender systems. The paper will provide substantial discussions of the notions of "normative" by cutting across history and philosophy of science, legal, and critical theory, as well as "algorithmics," and by confronting our studies led in engineering laboratories with critical algorithm studies.

Keywords

Machine learning, technical normativity, socio-technical normativity, Gilbert Simondon, neural networks, behavioral normativity

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

Introduction

It is generally agreed upon that "algorithmic systems" implementing machine learning techniques have significant normative effects upon the ways items are recommended and consumed, the ways choices are taken and justified. However, the aspect and extent of those "normative effects" are subject to much disagreement. Some claim that "algorithmic systems" significantly affect the way norms are conceived and instituted. Others maintain that algorithmic systems are actually black boxing more traditional normative processes (Diakopoulos, 2013). The problem partly lies in accounting for the different kinds of "normativities" that are, and can be attributed to, "algorithmic systems." Our paper proposes a substantial discussion of "algorithmic normativities" by accounting for the practices we witnessed in diverse engineering laboratories and by confronting literature in sociology, history and philosophy of technology (see for instance Beer 2017 and Ziewitz 2016). Before going any further let us begin by providing working definitions of "algorithm" and "norm," the value of which should be measured by the conceptual and practical relations they give rise to

[1]Université Libre de Bruxelles, Centre de Théorie Politique, Belgium
[2]Université de Namur, Research Center in Information, Law and Society (CRIDS), Belgium

Corresponding author:
Tyler Reigeluth, Universite Libre de Bruxelles—Campus du Solbosch, Bruxelles 1050, Belgium.
Email: tyler.reigeluth@ulb.ac.be

when thinking about, or interacting with, algorithmic systems.

From an engineering standpoint, algorithms can be helpfully approached as stabilized and formalized computational techniques that scientists and engineers recognize and manipulate as an object throughout a variety of disparate *inscriptions* (e.g., formalized expressions, diagrammatic representations, coded instructions, traces of executions) and through a variety of disparate *actions* (e.g., formal proofs, intuitive manipulations, material implementations, concrete experimentations). Computational techniques only become algorithms when calculators find it useful to stabilize and formalize them (see Hill, 2016). Two further features must be added to make sense of what engineers commonly call ''algorithms.'' First, algorithms are, for most computer scientists, objects about which knowledge can be produced and communicated. Second, algorithms are, for most software engineers, things likely to be recognized in many different systems. In other words, to warrant the qualification of algorithm, a stabilized and formalized computational technique needs to possess, and be invested with, a certain significance within a certain practice (Pickering, 1995).

From a generic standpoint, vital, social or technical norms can be approached either as a *description* of a range of events, or as a *prescription* for a range of actions—it has convincingly been argued that, in most cases, these descriptive and prescriptive meanings cannot be satisfactorily disentangled (Canguilhem, 1943; Putnam, 2002). Three features should be emphasized. First, a norm expresses a relation between a set of disparate changes—social norms, for instance, typically describe or prescribe ways of behaving within specific situations. Second, a norm reveals itself when individuals are propelled to contrast their ways of behaving with others'—again, social norms typically manifest themselves whenever individuals experience specific situations as being problematic or conflictual. Third, a norm is both *instituted* and *followed*—social norms are typically instituted as the solution to a collective experience of a problematic or conflictual situation. The notions of ''normativity'' and ''normalization,'' to which we will hereafter refer, denote the processes through which an activity, respectively, comes to institute or to follow specific norms.

The following pages further investigate how distinct kinds of normativities come into play through these socio-technical assemblages we call algorithmic systems.

- The first section exposes some of the minute actions through which engineers willingly or unwittingly infuse technical systems with social norms. The design of a surveillance system will serve to unfold the ongoing mediations (e.g., collecting data, defining metrics, establishing functions) that allow engineers to normalize the behaviors of algorithmic systems (e.g., how algorithms improve their ability to discriminate between threatening and nonthreatening behaviors).
- The second section shows how successive changes in algorithms' structures and operations reveal distinct kinds of *technical* normativities—be they schematic, formal or material. A liminal genealogy of artificial neural networks shall provide a powerful illustration of this dynamic by engaging with their concrete functioning as well as their unsuspected potentialities (e.g., how the invention of optimization algorithms or parallel processors considerably transformed these networks).
- The third section proposes to conceive of learning machines as exhibiting a genuine social or *behavioral* normativity, insofar as their activities can be seen as exhibiting both ''norm following'' and ''norm instituting'' processes. The case of recommender systems shall allow us to identify markers around which these two processes seem to irremediably entangle (e.g., the intractability, the randomness, and the interactivity of algorithmic systems' behaviors).

The conceptual focus of the paper does not dispense us from indicating the empirical sources thanks to which we ground our developments. The surveillance system's empirical material derives mainly from ethnographic data collected between 2012–2015 and 2016–2019 while participating in two Research and Innovation projects (e.g., consortium meetings, interviews with engineers, laboratory visits). The neural network's empirical material is chiefly informed by readings of original technical literature (Rosenblatt, 1962), contemporary technical literature (Bishop, 2006; Mitchell, 1997) and published historical accounts (McCorduck, 1983; Crevier, 1993; Olazaran, 1993; Pickering, 2010). The recommender system's material freely draws upon knowledge acquired along an empirical study following a computational experiment on recommender systems conducted between 2016 and 2017, as well as of the consultation of other more traditional resources such as scientific literature and technical blogs.

Socio-technical normativity and surveillance system

The recent applications of machine learning techniques in assisted and automated decision-making raise hopes and concerns. On the one hand, there are hopes about the possibility to have an informational environment tailored to specific ends or the possibility to increase a firm's commercial revenues. On the other hand, there are concerns about the inconsiderate discriminations machine learning results might conceal or the unprecedented possibilities for governing ushered in by these techniques. Such concerns appear to be exacerbated by the patent difficulties to hold beings accountable, either because the responsibility of the decision is not attributable in any straightforward way, whether to an algorithmic system or to a moral person, or because the understanding of processes leading to case-specific decisions is threadbare (Burrell, 2016).

We propose to address these key issues by identifying *sites* that are crucial for engineers' representations and interventions as they negotiate encounters between social imperatives and technical constraints (see Bijker et al., 1987; McKenzie and Wajcman, 1985). Two such sites have been identified: *the metrics' definition* and *the databases' collection*.[1] Each respectively enables engineers to evaluate and design their algorithmic systems. For the most part, our claims are grounded in empirical inquiries we undertook over the last five years, within different research groups. The material collected consists of observations, discussions, interviews, and reports. Thus, our first task is to describe the actions through which engineers produce algorithmic systems following specific socio-technical norms as well as the actions through which engineers produce knowledge about their algorithms' behaviors. One case study in particular, the "Privacy Preserving Protection Perimeter Project," will illustrate the matters discussed.[2]

Funded by the European Commission between 2012 and 2015, this European "Research and Innovation" Project gathered a dozen private companies, research universities, and public institutions with the aim of designing a system able to automatically detect threatening behaviors. This rather abstract endeavour inevitably concealed multiple concrete technical challenges, the most predominant of which were the combination of visual and thermal cameras with microwave and acoustic sensors, as well as the development of robust and efficient detection, tracking and classification algorithms (i.e. working in real-time within uncontrolled environments). The consortium rapidly pinned down two "use cases" that would put them to work: a Swedish nuclear power plant (the Oskarshamns Kraftgrupp AB) and a Swedish nuclear waste storage facility (the Centralt mellanlager för använt kärnbränsle). The task impelled the engineers to undertake a socio-technical inquiry enabling them to spell out the various qualities with which users (i.e. the security workers) expected the system to be endowed.[3]

Understandably, the security workers expected the system to maintain the number of false alarms below a certain threshold—otherwise the system would end up *dis*organizing the perimeter surveillance instead of organizing it. The project's partners usually reframe the requirement in slightly more technical terms: they want to minimize the number of "false positives" (i.e. when a behavior is mistaken as threatening) and "false negatives" (i.e. when a behavior is mistaken as nonthreatening). The engineers then translate the socio-technical requirements into fairly simple and intuitive mathematical expressions, called metrics, that set standards for attributing numerical values to different aspects of the system's behaviors, aspects that are deemed particularly important and that can be empirically observed (Dewey, 1939; Porter, 1996). The matter thus lies in understanding how the human-based act of surveilling through categorization can be transformed into a machine-based problem of classification.

The very notions of "false positive" or "false negative" suppose that an algorithm-based classification can be compared to, and overturned by, a human-based classification. This human-based classification, acting as the reference norm against which all machine-based classifications are to be evaluated, is called the *ground truth dataset* (Jaton, 2017). The "dataset" part here refers to an ensemble of recorded scenes, each containing as many numerical sequences as there are available sensors (i.e. visual videos, thermal frames, sound recordings, and Doppler time-series). It provides the system with concrete instances of an abstract classification problem. These instances consist of actual scenes the algorithm must assign to its "correct" category (i.e. threatening or nonthreatening). The "ground truth" part refers to the categories or labels which humans—e.g. domain experts, computer scientists or Amazon turkers—have attributed to each sequence. It supplies the system with answers to the problem: the algorithm now has an external check for assessing the correctness of its classification.

How are these ground truth datasets concretely constituted? Constituting relevant datasets for algorithmic systems presents a genuine socio-technical challenge. Building on their socio-technical inquiry—which involves reading regulations, visiting sites, and interviewing workers—the engineers envisioned scenarios that would seek to encompass the range of threatening and inoffensive behaviors the system was expected to handle (e.g., a jogger running near the nuclear power plant, a boat rapidly approaching the waste storage facility, a truck driving perpendicular to one of the site's fences, etc.). The engineers then gathered on two occasions, each time for about a week, to enact and record the norms scripted in their scenarios (Figure 1). The scene itself is worth visualizing: engineers awkwardly running through an empty field sprinkled with cables and sensors, mimicking the threatening or inoffensive behaviors they want their algorithms to learn.

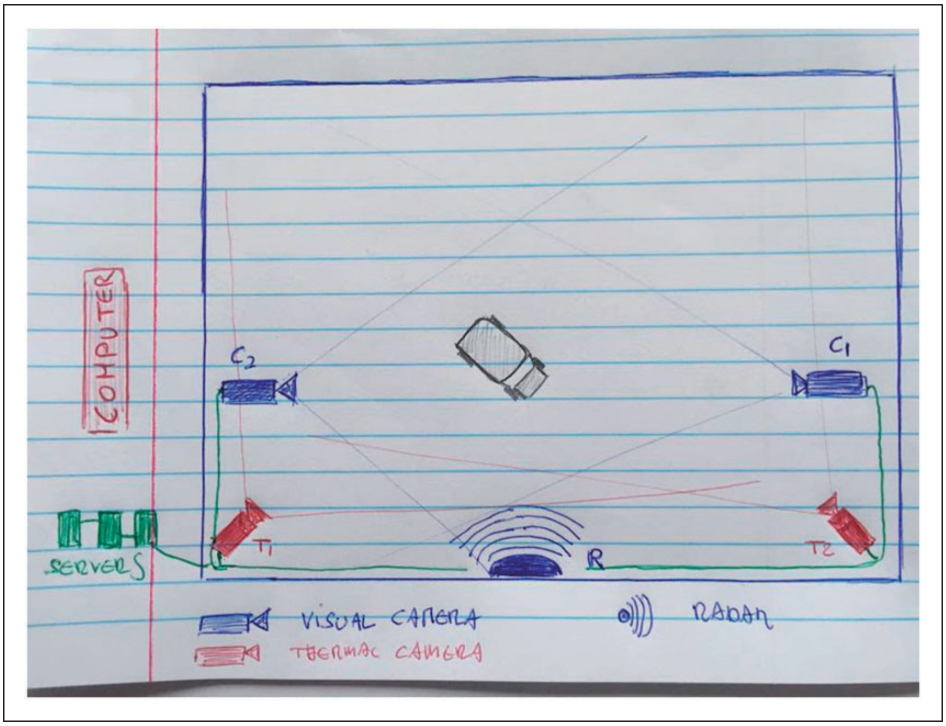


Figure 1. Personal sketch of the data collection field observed in United Kingdom in June 2014.

Returning to their laboratories, they spent a considerable amount of time annotating each frame of each sequence produced by each sensor, e.g. circling moving bodies, labeling their behaviors, etc.[4]

With their metrics defined and their data collected, engineers can, at last, *train* their algorithms, i.e. lead them to embody socio-technical norms, and *evaluate* them, i.e. measure relevant dimensions of their activities. The training process usually implemented first requires engineers to load and initialize the algorithm they wish to train (e.g., support-vector machine, random forest, artificial neural networks, etc.) on their laboratory's computer clusters. They then run a script which, as shown in Figure 2, (i) supplies the algorithm with a batch of scenes which algorithms are asked to classify (i.e. classification), (ii) compares the algorithm's guesses with the ground truth (e.g., evaluation), and (iii) modifies a number of the algorithm's parameters to improve the ways it reacts to specific videos (i.e. correction)—before iterating back to (i) until the training dataset is emptied.

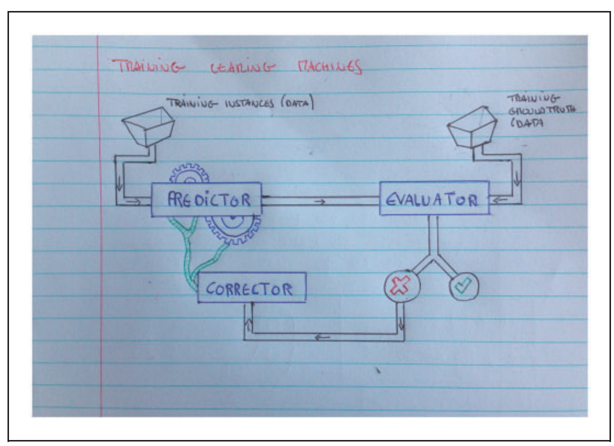


Figure 2. Personal sketch of a slide used by Robin Devooght on June 2017.

We have argued here—against growing claims that contemporary machine learning practices and techniques are progressively slipping through our hands—that approaches combining empirical and critical perspectives are likely to provide us with the means to productively engage with engineering practices, while also opening up possibilities of critique and intervention. The argument essentially rested on two claims. On the one hand, the ''metrics,'' which mathematically express the system's socio-technical norms, equip engineers with a way of assessing the comparative values of different algorithms. On the other hand, the ''ground truth dataset,'' which comprises the actual and possible instances the system is supposed to handle, provides engineers with local norms for telling the algorithm how to learn in specific instances. *In both cases, the norms are instituted by the engineers and are expected to be followed by the machines*. We believe that by identifying two distinct sites where engineers decisively mediate between social and technical constraints, we can productively transform the demands we place on algorithmic systems and provide robust indications as to where to intervene in critical cases in which normative conflicts and problems arise.[5]

Technical normativity and artificial neural networks

The recent craze surrounding ''machine'' or ''deep'' learning is usually explained in terms of ''available data'' (i.e. infrastructural changes made data collection accessible), ''algorithmic advances'' (i.e. technical changes made data processing possible) or ''economic promises'' (i.e. applications will make data particularly valuable). In focusing on the algorithmic advances, the following paragraphs propose to further explore three *kinds* of norms (imaginal, mathematical, and material) algorithms manifest whenever engineers attempt to endow them with specific capacities. The genealogy of artificial neural networks—a subset of ''machine learning'' techniques whose applications may be seen in surveillance systems (see ''Socio-technical normativity and surveillance system'' section) as well as recommender systems (see ''Behavioral normativity and collaborative filtering'' section)—will shed light on the *technical* norms algorithms impose on engineering practices. Before unfolding the technical transformations artificial neural networks underwent, we will briefly sketch their inception.

The earliest trace of the artificial neural network scheme can be found in a technical report authored by the American psychologist Frank Rosenblatt (Rosenblatt, 1962) when based at the Cornell Aeronautical Laboratory. By that time, several researchers were interested in exploring the nature of learning processes with the embryonic tools of automata theories. Distinctively, Rosenblatt decided to address the problem of learning processes by both investigating neurophysiological systems *and* constructing intelligent machines. Artificial neural networks would eventually emerge from these projects exploring ''natural'' and ''artificial'' learning, as the blueprint of a machine capable of ''perceiving'' and ''recognizing'' visual patterns, as a machine capable of ''learning'' to differentiate between geometrical forms. Thus, a neural image of intelligence gradually turned into concrete technical objects that would later be programmed for the IBM 704 and hardwired as the MARK I.[6]

It has been convincingly argued that the singularity of technical objects is best grasped through the schemes describing their operations within different environments, rather than the uses to which they are subject or the practicalities of their actualization (Simondon, 1958: 19–82). Thus, the singularity of artificial neural networks lies in their technical scheme, rather than the

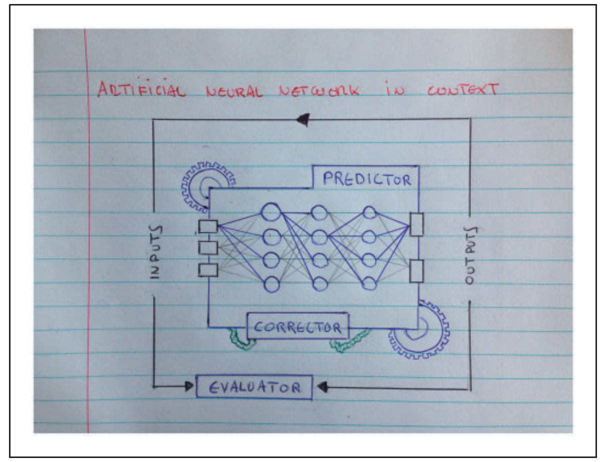


Figure 3. Diagram of artificial neural network within an environment.

general regression, classification or clustering purposes they serve, or their specific implementations in Python's Theano or TensorFlow libraries. A technical scheme describes the *parts* composing the technical object and relates their *operations* within the technical object's functioning (Polanyi, 1966: 38–40). Most schemes are constructed from material traces (objects, descriptions, diagrams, etc.) by technicians or historians interested in thinking about, and intervening upon, specific technical objects.

In the case of artificial neural networks, the scheme always couples the algorithm to an environment (Figure 3). In Rosenblatt's design, light sensors allow the network to capture signals from its environment that end up being classified and light emitters enable the network to display the results of its classifications. Crucially, Rosenblatt articulated the artificial neural network and its environment with a genuine feedback mechanism, which would work as an experimental controller for correcting the emitter's output until it produced the desired response.

The artificial neural network's scheme consists of two main parts: neural units and synaptic edges (see Figure 4). Each neural unit *receives* an incoming signal and *produces* an outcoming value—the neurons

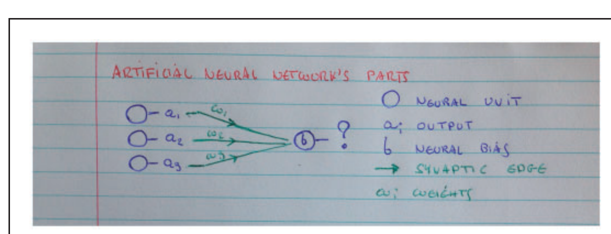


Figure 4. Diagram of the parts composing artificial neural networks.

generally contain numerical parameters (i.e. the bias) and share an activation function. Each synaptic edge connects two separate neural units—the synapses generally contain numerical parameters (i.e. the weight). The question now is how these parts interact.

The first operation is prediction. It typically moves from left to right and leaves the network unchanged (see Figure 5). A cascade of operations combines the input signal and the network's parameters (i.e. synapses' weights and neurons' biases) in order to produce specific output values—depending on the problem, the signal is classified as "triangle" or "circle," as "threatening" or "nonthreatening," etc.

The second operation is learning. It typically moves from right to left and alters the network's parameters (see Figure 6). A cascade of operations compares the outputs' values (i.e. the datasequence the algorithm classified) and the ground truth information (i.e. the sequence data the engineer annotated) in order to update and correct the network's parameters (i.e. synapses weights and neurons' biases).

More than a mere description encompassing different but related technical objects, this technical scheme opens a field of possible manipulations for engineers, ranging from minor modifications (How many neurons? How many synapses? etc.) to major modifications, which lead to new lineages of artificial neural networks (What if the neurons' activation function changes? What if peculiar synaptic connections are allowed? etc.). Thus, technical schemes are always both objectal and imaginal (Beaubois, 2015). If technical objects "have a life of their own," to borrow Ian Hacking's (1983: 262–275) words, the problem then lies in accounting for their schemes' becoming, as it mutates through but also resists, successive inventions (Leroi-Gourhan, 1945; Simondon, 1958: 19–82). The following paragraphs briefly sketch two other episodes neural networks went through between the early 1950s and the late 2000s, further illustrating this dynamic of invention and resistance.

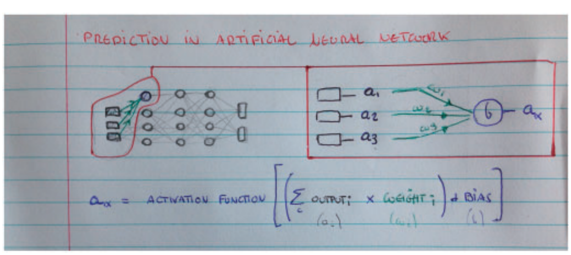


Figure 5. Diagram of the prediction's operation in artificial neural networks.

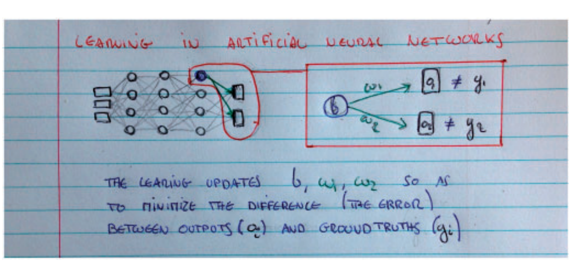


Figure 6. Diagram of the learning's operation in artificial neural networks.

The experimental, mathematical, and commercial successes of artificial neural networks rested upon their ability to adjust synapses' weights and neurons' biases in order to recognize incoming patterns—i.e. to learn to classify. However, the limitations of early artificial neural networks rapidly became apparent and problematic (see Olazaran, 1993: 347–350). The single-layer artificial neural networks proved incapable of solving important families of problems, e.g. they could not learn to recognize resized or disconnected geometrical patterns (Rosenblatt, 1962: 67–70). On the other hand, learning capacities of multilayer artificial neural networks appeared to depend more on engineering skills than on their intrinsic qualities, e.g. there existed no procedure guaranteeing that the learning would converge (see Olazaran, 1993: 396–406). These combined limitations to algorithmic learning capacities are traditionally seen as the onset of a significant drop in financial and scientific interest in artificial neural networks and artificial intelligence.

It was not before the mid-1980s that a satisfying learning rule, named "backpropagation," would be identified for multilayer artificial neural networks—more or less simultaneously and independently by Paul Werbos, David Rumelhart, and Yann Le Cunn (Olazaran, 1993: 390–396). In a nutshell, the learning algorithm measures the difference between the network's and the environment's responses (i.e. error function); it then computes the relative contribution of each synapse and neuron to the measured error (e.g., chain rule's partial derivatives); and finally updates the synapses' weights and neurons' biases so as to reduce the overall error. The learning problem is thus recast as an optimization problem, in which the main objective is to minimize classification errors. Backpropagation's mathematical scheme significantly extended artificial neural networks' problem-solving capacities.

Thereafter, in the late 1980s multilayer artificial neural networks gave rise to interesting applications in areas such as natural language processing, handwriting recognition, etc. (Olazaran, 1993: 405–411). However, the computational resources required during the learning phases quickly presented a significant impediment: depending on the number of layers, on the size of the datasets and on the power of the processors, it could take up to several weeks to find weights and biases minimizing classification errors. Thus, the applications using artificial neural networks did not stand out from other machine learning algorithms (e.g., "k-nearest neighbors," "random forests," "support vector machines," "matrix factorization"). Their more recent proliferation—often referred to as "deep learning"—appears to be intimately tied to advent and generalization of massively parallel processing units.

In the late 1990s, Nvidia Corporation, one of the largest hardware developers and manufacturers for video games, equipped their graphic cards with specific-purpose processors. Contemporary video games typically involve large-scale mathematical transformations that can be performed in parallel, e.g. the changing values of the millions of pixels displayed on our screens are usually processed simultaneously, instead of being calculated sequentially. In the late 2000s, researchers in computational statistics and machine learning revealed the mathematical affinities shared by the graphical operations in video games and the learning operations in pattern recognition. Indeed, both involved countless matricial calculations that could be distributed over several processing cores and executed independently. This rapid outline indicates the extent to which material developments brought forth or actualized latent algorithmic capacities.

Thus far, our approach has emphasized three kinds of technical norms algorithms impose on engineering practices. The first has to do with the algorithm's imaginal affordances in relation to the development of technical systems—in this case, a single abstract representation of neural processes brought about multiple lineages of algorithms. The second deals with the mathematical properties of the algorithm's structures and operations—in this case, the invention of a convergent algorithm significantly extended the abilities of multilayer artificial neural networks. The third bears on the interplay between material constraints and computational possibilities—in this case, the advent of Graphical Processing Units transformed the range of problems algorithms can solve. Thus, the genealogy of artificial neural networks led us not only to a concrete account of algorithmic processes, but more importantly still, to a comprehensive understanding of the imaginal, mathematical and material dynamics that drive their technical becoming.

Behavioral normativity and collaborative filtering

It is generally observed that computers blur an entrenched ontological dichotomy between two kinds of beings: living organisms and automated machines (see Fox-Keller, 2008; 2009). Kant's far-reaching conception of self-organization famously epitomizes the matter at stake: organisms' norms are thought to be intrinsic to their activities—organisms are, in this sense, "norm-instituting" beings—and machines' norms are thought to be extrinsic to their activities—-machines are, in this sense, "norm-following" beings (Barandiaran and Egbert, 2014). The most convincing conceptual markers grounding this longstanding dichotomy are generally to be found in the widely shared reluctance to attribute "problems" and "errors" to behaviors exhibited by machines, computers, and algorithms (Bates, 2014). Thus, technical errors are usually thought to be *rewritable* either as engineering *failures* or socio-technical *problems* (see, respectively, Canguilhem, 1943; Turing, 1948).

Contemporary algorithmic systems appear to further complicate any a priori partitions between distinctively organic and machinic behaviors. Indeed, *learning* machines are, to a certain extent, capable of modifying their structures so as to respond to modifications in their milieu without any specific human interventions. In this regard, they are often seen as exhibiting a form of behavioral *plasticity*, which was long held to be distinctive of organic activity. The minute operations followed by learning machines and the effects they produce, currently remain beyond satisfactory understandings. As such, there is an aspect of what they do that is *intractable*, that lies beyond our current capacities of prediction and control. This section further investigates these *underdetermined* behaviors and suggests that both the human *and* algorithmic aspects of machine learning systems need to be approached as genuinely behaving, that is as performing together norm-following and norm-instituting activities.

The case of collaborative filtering recommender systems deployed on commercial platforms such as YouTube or Netflix can help us ground these considerations (Seaver, 2018). In a nutshell, these systems seek to foster unprecedented interactions between items and users by looking at past interactions between similar items and users. Practically speaking, the *metrics* deployed for evaluating the performances of such systems generally attempt to measure different aspects of user engagement and the *data* collected for training these algorithmic models largely consists of user interaction histories. The concrete processes, which are to a large extent unscripted, remain difficult to explain, even retrospectively, for the engineers who designed the algorithms. The dynamics of imitation and variation performed on these platforms need to be conceived, we argue, as a *social activity*, involving an interactive and iterative process between the users' and the algorithms' behaviors.[7]

The distinction between norm-instituting and norm-following can be further understood as a difference between two processes of determination. Most technical entities, it has been argued, exhibit a relative ‘‘margin of indetermination’’ that makes them more or less receptive to ‘‘external information’’ and responsive in terms of ‘‘internal transformations’’ (see Simondon, 1958: 134–152), e.g. the progressive wear and tear of a bolt connecting metal parts allows for their mutual adjustment, an engine's governor constantly regulates the train's speed despite changes of load and pressure, a warehouse logistically adapts to a changing book order, a recommender system responds to newly received interactions. The distinction between norm-instituting and norm-following can thus be reframed as a distinction between two kinds of determination: a determination will be said ‘‘convergent’’ or ‘‘divergent’’ depending on whether it *restricts* or *expands* the behavioral variability of a technical entity. Both empirically and conceptually speaking, the difficulty thus lies in being able to account for the divergent determinations of certain machine learning processes.

In this light, the singularity of machine learning applications should be understood in terms of the relative significance of the margin of indetermination exhibited. Indeed, in most machine learning applications, the behavior of the algorithmic system (i.e. the specific predictions displayed) may be periodically transformed (i.e. the values of the model are updated) depending on the relevant incoming information (i.e. particular sets of data). By and large, engineering work is dedicated to restraining the recommender system's variations—instituting metrics, defining error functions, cleaning datasets—and aligning them with the companies' interests. On the one hand, the determinations of learning may be seen as converging toward rather specific pre-instituted norms. However, the actual learning process, as Adrian Mackenzie (2018: 82) has argued, has more to do with stochastic function-finding than with deterministic function-execution—the resulting model corresponds to one among many possible configurations (see also Mackenzie 2015). Thus, and on the other hand, the concrete determinations of the system's behavior can hardly be seen as converging toward any pre-instituted norms.[8]

Let us take a closer look at the actual processes that unfold when producing recommendations. Collaborative filtering techniques typically seek to achieve a delicate balance between sameness (always recommending similar items) and difference (only recommending different items) solely by looking at and drawing on patterns of interactions between users and items. The collaborative machines—that may well be instantiated by many different techniques, such as k-nearest neighbors, matrix factorization, neural networks—are usually interpreted by recommendation engineers as capable of producing models that represent *actual* relations and suggest *potential* relations between users and items. More concretely, each learning step supposes that the *error function* (also called loss or objective function) measures the recommendation's quality and that the correction algorithm (e.g., backpropagation) adjusts the model's parameters accordingly. We can now, with these developments in mind, rework the problem with which we opened the section: error as a marker for thinking the human–machine distinction.

If, indeed, the very process of learning rests on the possibility of erring (Vygotsky, 1978), we propose, following David Bates (Bates and Bassiri, 2016; see also Malabou, 2017), to seize both entangled senses in which erring may be understood: erring as ‘‘behaving in unexpected ways’’ (i.e. errancy), and erring as ‘‘susceptible of being corrected’’ (i.e. error). In both cases, erring supposes that a relation between an individual and an environment be experienced as problematic, that the norms instantiated within behavioral performances are experienced as only certain ones among many possible others. The passage from error to errancy depends on recognizing social activity as open-ended enough for error to appear not only as the mere negation of a norm, but rather as the affirmation of a different norm. We would like to suggest, here, that a more cautious look at learning processes could help us frame these algorithmic systems *as erring within a broader social activity*, in this instance: the algorithmic production of cultural tastes.

Two remarks should suffice here to pin down what is at stake. First, the overall learning process is open-ended: although each learning step stops when the model's parameters minimize error and overfitting, the overall learning process consists of an indefinite sequence of learning steps, which are periodically picked up again once there is enough fresh data from incoming interactions. Second, the recommendations are ‘‘moving targets’’ (Hacking, 1986): the learning process is shaped by interactions as much as it shapes interactions, i.e. recommender systems learn *and* produce interactions. We therefore contend that the iterativity and interactivity of recommending activities are grounds for conceiving classification operations performed by recommender systems as targeting both certain user behaviors as well as certain algorithmic predictions. The dynamic of these ‘‘looping kinds’’ (Hacking, 1986), brings recommendation, as a social activity, closer to errancy than error. Indeed, what would it even mean from this perspective to produce a false recommendation?

The algorithmic system can therefore be seen as genuinely *behaving*, that is as performing an algorithmic *activity* exhibiting its own form of normativity. The behavioral normativity, once the notion of behavior is released from individualized moorings and reconnected with the social activity within which it unfolds, can be productively conceived as a dynamic of norm following *and* norm institution—rather than the instantiation of social norms within technical ensembles (i.e. socio-technical normativity) or the instantiation of technical norms through acts of invention (i.e. technical normativity). In this specific case, the recommender system needs to be conceived as *a* social partner behaving within a determined social activity: their behaviors bear social significance and affect the value of other behaviors.

Thus, the notion of behavioral normativity leads us to reconsider how the redistributions of norm-following and norm-instituting behaviors between humans and machines actively reshape social activities. We might ask then: "What is learning and where is learning occurring within the behavioral distributions of a given activity?" In doing so, the primary *focus* or *locus* of analysis is productively displaced toward the unfolding of social activities, thereby indefinitely deferring the quest to reveal the "true norms" that inform algorithmic systems (e.g., "we are *manipulated* by algorithms") or social systems (e.g., "algorithms are *biased*"). We can now better attend to what it means and why it matters for machine learning algorithms to behave in unforeseen and unexpected ways, thus opening the prospect that they become sites of normative invention. We hope, in this way, to have contributed to the concerted efforts that need to be made to integrate algorithmic behaviors into the field of action it analyzes by offering conceptual and methodological tools for claiming their effects. Our contention is that one way of addressing this challenge lies in showing how this normativity makes sense within culture.

The tension experienced by most people when confronted with technical systems can, in part, be understood in terms of the difficulties they have in experiencing and making sense of how, where, and when machine behaviors perform with ours. The cultural images of technical systems have traditionally allowed us to stabilize the perceptive and motor anticipations determining our relationships with them, e.g. the individual body performing with a simple tool or the industrial ensemble organizing human and machine labor (Simondon, 1965-1966; Leroi-Gourhan, 1965). The problems, in the case of learning machines, are tied to the variable distributions of human and machines behaviors (Collins, 1990: 14) and to the unfamiliar ability algorithmic systems have of instituting norms. The tension can thus in part be reframed as the problem of understanding these behaviors in relation to the social activities within which they unfold and that they take part in shaping.

The concept of behavioral normativity foregrounds the importance of social activity and the afferent margin of indetermination it allows behaviors to inform. If there is no room left for error, then all behaviors that do not execute or follow the established, programmed norm will be disregarded or repressed as useless or inefficient. If, on the other hand, those behaviors that do not perform as expected are attributed value and attended to, they can participate in instituting a new norm-following dynamic. Thereby, machine learning forces us to reconsider long standing divides between machines' and organisms' behaviors, between those behaviors that repeat and those that invent norms. Our conceptual proposition invites alternative and more demanding normative expectations to bear on engineering and design practices whereby the margin of indetermination of an algorithmic system would be increased rather than reduced.

Conclusion

The overall aim of this paper was to approach algorithmic normativities in a different light, with different questions in mind, with different norms in sight.

The first investigation into socio-technical normativity showed how engineering practices come to stabilize and embed norms within certain systems by setting up plans or programs for learning machines to execute (metrics, ground truth dataset, optimization function). The socio-technical perspective, although a necessary starting point, is insufficient if taken in and of itself. Indeed, it demands that we be able to define what is "social" and what is "technical" within a given system, and how their given normativities come to be entangled. In this light, we proposed to qualify social and technical normativities in terms of operations they perform, rather than properties they possess.

This led us to look at different algorithmic systems from the standpoints both of their *technical* operations and their social or *behavioral* activities. The section on notion of technical normativity sketched the mutations of an algorithm's technical scheme and exposed how its norms both induce and resist invention, thereby granting technical objects imaginal, mathematical or material consistencies. The final section on behavioral normativity allowed us to consider certain conditions (indetermination, divergence, errancy) under which an algorithm can be seen as taking part in the norm-following and norm-instituting dynamics that characterize social activities.

This normative pluralism can help understand how contemporary algorithmic systems, comprised of multiple structures and operations, simultaneously fulfill engineering aims, express technical resistances and participate in ongoing social processes. Generally speaking we can advance that:

- the aims engineers pursue always depend upon certain algorithmic capacities, in this sense algorithmic techniques normalize engineering practices (i.e. the objective function always depends on an efficient learning rule);
- the socio-technical system's behavior is subject to engineering aims that normalize its learning activity (i.e. the metrics and the ground truth datasets determine what counts as an error);
- the algorithmic system's learning processes unfold genuine norm instituting behaviors (i.e. the system's outputs periodically affect the very activities that have to be learned).

The value of our approach lies in its ability to provide a nuanced account of what is too often presented as *one* opaque, impenetrable or ethereal *system*. Algorithmic system*s* must instead be seen, we argue, as inhabited by *normative tensions*—between technical, socio-technical, and behavioral normativities. We sketched this pluralism with specific practical and theoretical problems in mind. No doubt others would be led to explore different normativities pervading algorithmic systems. Our effort has largely consisted in showing that behind each "norm-following" instance there is an institution, and conversely, that every institution requires its norms to be performed, to be played out, even at the risk of them being transformed by their very performance.

Acknowledgement

We would like to thank our friends and colleagues Nathan Devos and François Thoreau for discussions and criticisms, the two anonymous reviewers that made very productive critiques and suggestions, as well as the editors of this issue, Francis Lee and Lotta Lotta Björklund Larsen.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This paper was in part made possible thanks to the support of the Belgian FRS-FNRS funded research project "Algorithmic governmentality" which sponsored both Jérémy Grosman's and Tyler Reigeluth's PhD research. In addition, J. Grosman benifited from H2020 funding made available by the European Commission for the "Privacy Preserving Protection Perimeter Project" and "Pervasive and User Focused Biometrics Border Project".

ORCID iD

Tyler Reigeluth https://orcid.org/0000-0002-9908-1052

Notes

1. We do not claim that these sites are the *only* sites where social and technical normativity are entangled—one could also consider: (i) the choice of the relevant features (i.e. that select and neglect what counts as a relevant information, such as speed, trajectories or colors); (ii) the choice of the objective function (that determine the learning by providing a computational metric); (iii) the choice of the machine learning algorithm (that constrains the kind of things to be learned, such as sets or series).
2. The two inquiries, particularly relevant to the argument presented here, were led in the course of another two European "Research and Innovation" initiative, respectively, titled "Privacy Preserving Protection Perimeter Project" (2013–2016) and "Pervasive and User Focused Biometrics Border Project" (2016–2019). The data collected, during these projects, consisted of minutes of consortium meeting, attendance to demonstrations, as well as 30 semi-directed interviews with engineering partners. Grosman J (2018) Ethical and social issues, "Technical Report in the frame Pervasive and User Focused Biometrics Border Project" (PROTECT). The most readable general introduction to machine learning is undoubtedly Mitchell (1997). Bishop's (2006) textbook is also an invaluable resource. The course given by Andrew Ng (2011) at Stanford is an equally good place to start.
3. To give a more concrete sense of the specific challenges that need to be overcome: the system needed to remain indifferent to weather variations (e.g., Swedish snow, fog, and sun) as well as to the surrounding fauna (e.g., squirrels, rabbits, foxes, moose, etc.).
4. The "data collection" happens to be of tremendous importance not only, from the perspective of the power plant and waste facility's staffs, for getting the system to work correctly, but also, from the perspective of engineers working in computer vision, for getting standard datasets against which to compare their algorithms—mainly: providing benchmarks and organizing competitions.
5. Considering just one short example showing some consequences of our approach to the so-called opacity and unaccountability of algorithmic systems, the problems of algorithmic discrimination can be constructively reappraised as the problems of (i) characterizing the suspected social discrimination, (ii) constructing metrics likely of measuring such discrimination (and ruling out inconsiderate algorithms), and (iii) spelling out requirements that datasets must meet to avoid such discriminative consequences.
6. The empirical material concerning the genesis of neural networks builds on an extensive reading of the existing historical literature: McCurdock (1983), Olazaran (1993), Kay (2001), Pickering (2010), Cardon et al. (2018) as well

as selected readings of the primary literature: Rosenblatt (1962). The contemporary literature provides valuable technical details, see for example the short presentation in Mitchell (1997), the more substantial treatment in Bishop (1995) and the online course given by Geoffrey Hinton for Toronto University dedicated to neural network and available on the Coursera platform in 2012. For a more general overview of the period, see for instance Heims (1991) and Edwards (1996).

7. The most relevant textbooks are Francesco Ricci, Lior Rokach, Bracha Shapira and Paul B. Kantor (2011) and Aggarwal (2016). The annual proceedings of the ACM conference on *Recommender Systems* provide an invaluable overview of the public and private research conducted in recommender systems. Netflix's or Spotify's technical blogs, as well as the less official blogs of their employees, give invaluable insights into recommendation practices. The account is also informed by an empirical inquiry, led between 2016 and 2018, during which we documented a series of computational experiments attempting to uncover the capacities of recurrent neural networks for predicting sequence of interactions for the offline version of a collaborative filtering based recommender problem.
8. This argument rests on the rather reasonable assumption that the empirical learning of machines cannot be completely reduced to the mathematical convergence of a function, the role of which is to find the extremal values of another function—minimization in the case of what is called error or loss function, maximization in the case of what is called objective function.

References

Aggarwal CC (2016) *Recommender Systems: The Textbook*. New York: Springer.

Barandiaran X and Egbert M (2014) Norm-establishing and norm-following. *Artificial Life* 20(1): 5–28.

Bates D (2014) Unity, plasticity, catastrophe: Order and pathology in the cybernetic era. In: Killen A and Lebovic N (eds) *Catastrophe: History and Theory of an Operative Concept*. Berlin: De Gruyter, pp. 33–54.

Bates D and Bassiri N (2016) *Plasticity and Pathology. On the Formation of the Neural Subject*. New York: Fordham University Press.

Beaubois V (2015) Un schématisme pratique de l'imagination. *Appareil*, 16.

Beer D (ed.) (2017) *Information, Communication and Society* special issue: The Social Power of Algorithms 20(1).

Bijker WE, Hughes T and Pinch TJ (1987) *The Social Construction of Socio-Technical Systems: New Directions in the Sociology and History of Technology*. Cambridge: Massachusetts Institute of Technology Press.

Bishop CM (1995) *Neural Network for Pattern Recognition*. Oxford: Oxford University Press.

Bishop CM (2006) *Pattern Recognitions and Machine Learning*. New York: Springer.

Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 1–12.

Canguilhem G (1943) *Essai sur quelques problèmes concernant le normal et le pathologique*. For the English translation, see Canguilhem G (1991) *The Normal and the Pathological*. Translated by Fawcett CR in collaboration with Cohen RS, New York: Zone Books.

Cardon D, Cointet J and Mazières A (2018) La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux* 211(5): 173–220.

Crevier D (1993) *AI: The Tumultuous History Of The Search For Artificial Intelligence*. New York: Basic Books.

Collins H (1990) *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge: Massachusetts University Press.

Dewey J (1939) *Theory of Valuation*. International Encyclopedia of Unified Science (ed. Neurath O) 2(4). Chicago, IL: Chicago University Press.

Diakopoulos N (2013) *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*. New York: Tow Center for Digital Journalism, Columbia University.

Edwards P (1996) *The Close World: Computers and the Politics of Discourse in Cold War America*. Cambridge: Massachusetts Institute of Technology Press.

Fox-Keller E (2008) Organisms, machines, and thunderstorms: A history of self-organization part one. *Historical Studies of Natural Science* 38(1): 45–75.

Fox-Keller E (2009) Organisms, machines, and thunderstorms: A history of self-organization part two. *Historical Studies of Natural Science* 39(1): 1–31.

Grosman J (2018) Ethical and social issues, "Technical Report in the frame Pervasive and User Focused Biometrics Border Project" (PROTECT).

Hacking I (1983) *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hacking I (1986) Making up people. In: Heller TC and Brooke-Rose C (eds) *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*. Palo Alto, CA: Stanford University Press, pp. 222–236.

Hacking I (1999) *The Social Construction of What: Introductory Topics in the Philosophy of Natural Science*. Cambridge, MA: Harvard University Press.

Heims S (1991) *The Cybernetics Group*. Cambridge: Massachusetts Institute of Technology Press.

Hill RK (2016) What an algorithm is? *Philosophy and Technology* 29(1): 36–59.

Hinton G (2012) Neural networks and machine learning. *Coursera*.

Jaton F (2017) We get the algorithms of our ground truths: Designing referential databases in digital image processing. *Social Studies of Science* 46(7): 811–840.

Kay L (2001) From logical neurons to poetic embodiments of mind: Warren S. McCulloch Project in Neuroscience. *Science in Context* 14(14): 591–614.

Leroi-Gourhan A (1945) *Milieu et Techniques*. Paris: Albin Michel.

Leroi-Gourhan A (1965) *Le Geste et la Parole II: La Mémoire et les Rythmes*. Paris: Albin Michel.

McCorduck P (1983) *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. Natick, MA: A. K. Peters Ltd.

Mackenzie A (2015) The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.

Mackenzie A (2018) *Machine Learners: Archaeology of a Data Practice*. Cambridge: Massachusetts University Press.

McKenzie D and Wajcman J (1985) *The Social Shaping of Technology*. Buckingham: Open University Press.

Malabou C (2017) *Métamorphoses de L'intelligence: Que Faire de Leur Cerveau Bleu?* Paris: Presses Universitaires de France.

Mitchell MT (1997) *Machine Learning*. New York: McGraw-Hill Education, 1995.

Olazaran M. A sociological history of the neural network controversy. *Advances in Computer Science* 37: 335–425.

Ng A (2011) Introduction to machine learning. *Coursera*..

Pickering A (1995) *The Mangle of Practice: Time, Agency and Science*. Chicago, IL: Chicago University Press.

Pickering A (2010) *The Cybernetic Brain: Sketches of Another Future*. Chicago, IL: Chicago University Press, 2010.

Polanyi M (1966) *The Tacit Dimension*. Chicago, IL: The University of Chicago Press.

Porter T (1996) *Trust in Numbers: The Pursuit of Objectivity in Public Life*. Princeton, NJ: Princeton University Press.

Putnam H (2002) *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press.

Ricci F, Rokach L, Shapira B, et al. (2011) *Recommender Systems Handbook*. New York: Springer.

Rosenblatt F (1962) *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. New York: Cornell Aeronautical Laboratory.

Seaver N (2018) Captivating algorithms: Recommender systems as traps. *Journal of Material Culture* 1–16.

Simondon G (1958) *Du Mode D'existence des Objets Techniques*. Paris: Editions Aubier-Montaigne [1989].

Simondon G(1965–1966) *Imagination et Invention (1965–1966)*. Chatou: Editions de la Transparence [2008].

Turing A (1948) Intelligent machinery. In: Copeland J (ed.) *The Essential Turing*. Oxford: Oxford University Press [2004], pp. 395–432.

Vygotsky L (1978) *Mind in Society*. Cambridge, MA: Harvard University Press.

Ziewitz M (ed.) (2016) *Science, Technology and Human Value*, special issue: Governing Algorithms, 41(1).

Original Research Article

Big Data & Society
July–December 2019: 1–12
© The Author(s) 2019
DOI: 10.1177/2053951719867359
journals.sagepub.com/home/bds

Data ratcheting and data-driven organisational change in transport

Liam Heaphy

Abstract

This article explores the process by which intelligent transport system technologies have further advanced a data-driven culture in public transport and traffic control. Based on 12 interviews with transport engineers and field-work visits to three control rooms, it follows the implementation of Real-Time Passenger Information in Dublin and the various technologies on which it is dependent. It uses the concept of 'data ratcheting' to describe how a new data-driven rational order supplants a gradualist, conservative ethos, creating technological dependencies that pressure organisations to take control of their own data and curate accessibility to outside organisations. It is argued that the implementation of Real-Time Passenger Information forms part of a changing landscape of urban technologies as cities move from a phase of opening data silos and expanded communication across departments and with citizens towards one in which new streams of digital data are recognised for their value in stabilising novel forms of city administration.

Keywords

Intelligent transport systems, real-time information, smart city, Big Data, organisational change

Introduction

There is a quiet 'revolution' underway in the transport sector as intelligent transport systems (ITS) technologies are used to increase efficiencies and integrate urban functions, through an alliance of well-established transport technology companies in partnership with city engineers and technologists. ITS originates in the context of managing and coordinating road use, but interfaces with air, rail and water systems (Williams, 2008: 3). While discursively enveloped in the promise of the smart city in recent marketing programmes such as that of Dublin (Coletta et al., 2018a; Kitchin et al., 2018), the history of ITS extends back further over several decades with the development of CCTV-supplied traffic control rooms, junction controller software, and scheduling systems for public transport. Therefore, it provides a counter-example to the creation of new urban corporate districts branded as 'smart' (Wiig, 2019), or reductive concepts of a universal urban science (Shelton, 2017), inasmuch as it accounts for a gradual evolution of 'smart' or intelligent technologies within a specific sector.

ITS forms the basis for a longitudinal study on how data-driven organisational control is being enacted in our cities, drawing attention to the concept of 'data-ratcheting' to show how these data-driven changes are iterative and leveraged off previous innovations. The evolution of ITS is dependent on long-standing commitments to shared infrastructure, and indicative of a shift to increasing autonomous management of public transport and traffic while overseen by human operators. The article seeks to contribute to the theoretical literature on standards and Big Data by building an explanatory account of data-driven organisational change based on a case study on the deployment of Real-Time Passenger Information (henceforth RTPI) in Dublin. It focuses primarily on the implementation

UCD School of Architecture, Planning and Environmental Policy, Dublin, Ireland

Corresponding author:
Liam Heaphy, UCD School of Architecture, Planning and Environmental Policy, Richview, Clonskeagh, Dublin 14, Ireland.
Email: liam.heaphy@ucd.ie

of RTPI on bus services and considers the wide array of technologies on which RTPI is reliant, including locational technologies, telecommunications and information standards.

The following section reviews existing theory on data infrastructures in relation to its bearing for understanding the deployment of new technologies. This is followed by a description of the fieldwork and methods underpinning the empirical content. The subsequent three empirical sections then detail the deployment of RTPI in Dublin, largely in temporal order and identifying the various phases of development. This covers earlier trials, the negotiation of common information standards, forms of data-driven behavioural or procedural change, and the consolidation of operational intelligence technologies. The penultimate section further discusses data ratcheting in relation to procedural and organisational change. The conclusion then discusses the evolution of data-driven change in transport and speculates on further research directions.

Combining studies of the mundane with data assemblages

The study of infrastructure has covered both epic transformations, such as the electrification of Western societies (Hughes, 1993) and the mundane 'boring things' (Star, 1999), such as the file folder system for computing and its consequences for systems management (Yates, 1993). In the former, Hughes' panoramic view of large technical systems repurposes the military term 'reverse salient', pertaining to line formation, to define where blockages in one domain may be cleared by progress in another. Infrastructure becomes a means of analysing how systems travel and extend into local contexts, and a basis for forming concepts to explore their many contingencies and the development of common standards of information exchange.

Infrastructure studies blend the archival patience of the historian with the attentive eye of the ethnographer and require a predilection for interrogating technical reports and manuals. In this vein, Edwards (2010) covers the pioneering work of climate observers in the 19th and 20th centuries and the process of reanalysis, where scientists reconcile data from multiple sources into uniform global datasets. This large temporal scale opens up concepts such as 'infrastructural inversion', where 'historical changes frequently ascribed to some spectacular product of an age are frequently more a feature of an infrastructure permitting the development of that product' (Star and Bowker, 2006: 233). It also facilitates the analysis of positive and negative externalities associated with standards, such as Meinel's historical study (2004) on the contingencies of the field of microbiology on particular visual and mechanical model kits.

The digital revolution has expanded the remit of infrastructure studies. Star and Ruhleder's (2001) seminal study on the creation of a unified, computer-based and networked worm system for scientists examines the frictions that occur around arcane choices of networking technologies or operating system, providing a useful distinction between first-, second- and third-order issues. First-order issues concern direct matter-of-fact phenomena and can be solved with existing resources and processes, such as finding and enabling a software option. Second-order issues reflect unforeseen contextual effects, such as the choice of one suite of procedures or standards over another and the path dependencies that result. Third-order issues are inherently political, involving ethical frameworks, theoretical paradigms and schools of thought. They may arise from combinations of first- and second-order issues, such as where an awareness emerges that legacy technologies are constraining accessibility, opening the path towards more fundamental questions on the collective good.

Following Kitchin (2014a), we can associate the centuries-old Big Data of climate science referred to above with *volume* and *variety*. However, it is with ITS and similar forms of sensor-fed, large-scale digital networks and models that we can see the increasing importance of *velocity* (i.e. real-time or close to it). ITS relies on complex assemblages of information networks, human engineers and controllers, transport policies, road-side sensors, in-vehicle computers and global positioning systems to choreograph in real-time an increasingly wide array of objects in space. It is part of a shift towards data-driven design and maintenance of urban transport systems, drawn into discourse on the 'smart city' through including 'smart travel' or transport in funding programmes and established practices of classifying smart technologies (Giffinger and Pichler-Milanović, 2007), as well as through the inroads into transport sought by data analytics companies.

Calls have been made to study actually existing 'smart cities' (Coletta et al., 2018a; Shelton et al., 2015), and engage with software by analysing critically how data interact with urban space and society (Kitchin and Dodge, 2011). This can take the form of following the data and 'attending to the sociotechnical fuzziness of data as it falls between epistemological problems, material infrastructure and organizational concerns' (Coletta et al., 2018b: 6). It implies looking to the 'proxies' of data, where data flows are transformed, bifurcated, or collated; whether that be errors and anomalies in sensor networks and transport systems (Reed, 2018; Reed and Webster, 2010), or the rollout of sophisticated city-wide projects in partnership with industry giants (Shapiro, 2018).

Furthermore, it requires attention to the power relations and values implied as these data are recruited into new modes of data-driven urban governance.

Shelton (2017) shows how Big Data-inspired visualisations of vacant lots may depoliticise ethnic and social inequalities, normalising the principles by which such inequalities are largely left in place. Consequently, it is useful to account for the totality of politics, data, values and materialities that underwrite such visualisations or policy tools. Kitchin and Lauriault (2014: 1) advance the concept of a 'data assemblage that encompasses all of the technological, political, social and economic apparatuses and elements that constitutes and frames the generation, circulation and deployment of data'. It functions as a means of expanding discussion to how data reshape society, and resultantly, how data-driven change leads to further developments and the creation of new path dependencies. This aids in drawing attention to the politics of Big Data (Shelton, 2017), the rights of citizens to the digital city (Foth et al., 2015), and data literacy (Gray et al., 2018).

This article takes its cue from data assemblages as a means of exploring how urban datafication reshapes urban management and control. Kitchin (2014b: 25) pursues data assemblages in relation to how data can usher in new regimes of data-driven societal control, tabling a broad array of constituent factors. The framing of RTPI as a data assemblage opens discussion into how 'each apparatus and their elements frame what is possible, desirable and expected of data', including governmentalities, materialities, practices and systems of thought. This article is particularly concerned with *procedure*; how transport systems have integrated data-driven technologies into their internal modes of organisation and coordination rather than their changing relationship to passengers and broader society.

Modern intelligent transport technologies, it is maintained, represent an instance of urban datafication and consolidation, as organisations experiment with Big Data in the interests of increased efficiency and performance. This process of urban datafication can be described with recourse to three translations. Firstly, *data expansionism* occurs through the deployment of sensing technologies and ICT infrastructure to create new datasets. Concepts from infrastructural studies such as 'reverse salients' help explain how data platforms come into being as new technologies facilitate change and transport authorities and providers build necessary infrastructure. Secondly, as new data sources become available, *data experimentalism* ensues as a range of actors compete to create new services and products. Finally, a third phase of *operational consolidation* can be delineated, as new forms of data-driven behavioural change, based on the dynamic treatment of real-time data, are encoded into procedures and organisational technologies. A continual negotiation and policing of standards is necessary to ensure data-driven processes can function to a required degree of resilience, for which the study of first-, second-, and third-order issues aid in our comprehension of how this occurs on different levels. Finally, the titular concept of *data ratcheting* functions as a creative interlacing of these final and overlapping phases as new functions, products and services are discovered and implemented in a context of data-driven rationalisation.

The fieldwork introduced below corresponds to recent phases of data expansionism and operational consolidation. However, the implementation of RTPI in Dublin, particularly on bus services, evidences the long infrastructural timelines of ITS. Therefore, the section thereafter covers this initial period from the 1970s until the near-present, before then considering how RTPI standards were negotiated and how data-driven organisational change ensued.

Fieldwork and methods

Dublin's RTPI system covers Bus Átha Cliath (henceforth Dublin Bus), Bus Éireann (a public national coach company), Luas (tram system), and Iarnród Éireann (the public national railway company). Dublin Bus registered 136.3 million journeys in 2017, as compared to 37.6m for the Luas, and 45.5m for Iarnród Éireann nationwide (NTA, 2018: 201). The focus of research reflects the operational area of Dublin Bus, corresponding to the Greater Dublin Area (GDA). The GDA comprises four predominantly urban local authorities with a combined population of 1,347,359 in the 2016 census and three surrounding rural counties. There is no corresponding transport authority for the GDA. Instead, the national scale tends to be the favoured tier for integrated services across local authorities (Coletta et al., 2018a: 4). The National Transport Authority (NTA) is responsible for both national and regional transport planning including the GDA.

The choice of the fieldwork site and the topic of transport technology reflect a wider objective of tracking Dublin's self-promotion as a 'smart city' as part of a large multi-researcher project. It forms one of several complementary case studies on how smart technologies and Big Data are transforming urban life (cf. Cardullo and Kitchin, 2018; Coletta et al., 2018a; Perng and Kitchin, 2016). The empirical research presented in this article draws on 12 semi-structured interviews derived from two related fieldwork datasets. The first six are a subset are drawn from a larger set of 77 interviews conducted with government and city workers, corporations, and other stakeholders on Dublin's emerging smart city strategy in 2015/2016. These include five transport engineers from local and national government and one transport consultant, all of whom discuss ITS in relation to traffic control, road design and maintenance, and public transport reform. The remaining six are part of subsequent in-depth studies with transport operators (Bus Éireann, Dublin Bus, Luas) and traffic control room engineers (Dublin City Council) on interrelated RTPI and traffic management technologies in 2016/2017.

Interviews were conducted in situ in back offices (NTA, Luas, Bus Éireann) and control rooms (Dublin Bus and the traffic control rooms of Dublin City Council and South Dublin County Council). Observational fieldnotes from visits to the control rooms and impromptu conservations with operators provided additional perspectives. Interviewees were asked to explain their roles, the deployment of RTPI in their organisation and its relationship to other technologies, decisions on standards and their implementation, and their coordination with other transport providers and regulators. The insights offered by participants were supplemented by reports and secondary literature to inform the longer timeline of ITS in Dublin.

Four decades of pilots in Dublin

In Dublin as elsewhere, the development of ITS can be seen to be dependent on a supporting infrastructure which allows innovation to occur over multi-decadal timescales. There were many instances of transport innovation over the last few decades, but without sufficient coordination between the relevant bodies, they failed to scale up or attract sufficient continuity support from national or local government.

RTPI allows public transport users to consult the predicted departure time of services and determine the most efficient mode of travel, based on the latest information on traffic delays, interruptions and capacity. RTPI improves perceived reliability as passengers rate their transport providers more highly if they are kept informed of how the system is performing and can make more sophisticated information-driven travel decisions (Caulfield and O'Mahony, 2009; Watkins et al., 2011). Although transport operators can fall back on scheduled timetables and rosters to provide basic transport infrastructure, the higher levels of service attained with transport technologies are critically dependent on software and automation.[1]

RTPI is dependent on automatic vehicle monitoring (AVM), a technology which reports real-time information on the location of vehicles back to a central server. AVM dates back to the late 1960s, with trials in the United States of America and other nations to develop radio-based triangulation systems to an acceptable degree of accuracy and temporal resolution. It was tested for select bus routes in urban areas throughout the 1970s with varying degrees of success (Roth, 1977). For mass transport systems, AVM promised a more efficient and less costly alternative to the 'point men', employed to record stop times of buses and inform schedule adherence and redesign (Roth, 1977). These elemental technologies of control contribute towards tackling the perennial issues of buses running ahead of schedule (understood to be considerably worse than running behind, as it causes disruption to drivers further back down the line and frustration to patrons) and maintaining even headway (where buses are distributed evenly along the route). It was recognised in the late 1970s that AVM could be used to provide both frequent and reliable service information to passengers and also integrate with traffic control systems to give signal priority to oncoming public transport vehicles running behind schedule (Symes, 1980: 237). Dublin Bus first trialled AVM in the 1970s using odometers fitted on buses that reported by radio to a central server every 45 seconds, subsequently rolled out to all bus depots by 1981. In 1985–1987, Dublin Bus also trialled traffic prioritisation based on a combination of infrared transponders installed on buses and roadside detectors. This was linked to the AVM system, but funding was not available to continue a successful pilot. The AVM system continued until the end of its useful life in the 1990s but was not replaced, with controllers reverting to radio contact with drivers (World Bank, 2011).

In 2001, a second trial of next-generation GPS-based Automatic Vehicle Location (AVL) called 'Q-time' was conducted on select Dublin Bus routes, and for the first time, RTPI signage was fitted (Caulfield and O'Mahony, 2004). This trial continued for three years during a period in which there was perennial threats (or opportunities) to liberalise the transport sector. The Irish Minister for Transport announced in 2002 that 25% of bus routes in the capital were to be opened to private competition, while also proposing the development of a 'Dublin Land Use and Transport Authority' in line with best practice across the European Union (Caulfield and O'Mahony, 2003: 2). Privatisation in transport is associated with market deficiencies including the dumping of non-profitable but socially important routes, and fare-creep on monopolised well-transited routes. Observing the UK experience, further inefficiencies may include competing companies serving the same routes, multiple and incompatible ticketing options, and the duplication of management and control resources (cf. Sørensen and Gudmundsson, 2010). Privatisation is therefore often accompanied with a parallel investment in new regulatory structures to mitigate these issues while proceeding on the ideological basis of lowering public investment costs and mitigating militant trade unionism (Gomez-Ibanez and Meyer, 2011).

In this context of uncertainty and privatisation in the early 2000s, Dublin Bus did not attain funding for a city-wide implementation of their Q-time pilot. Therefore, by the time RTPI was finally funded and implemented city-wide on buses from 2009 to 2011, it was a mature and largely consolidated technology commonplace in European cities like Gothenburg and Helsinki since the mid to late 1990s. Acting as a reverse salient, the arrival of GPS was making locational technologies part of everyday experience for both transport operators and personal devices (Kitchin, 2014b: 58), and could overcome technology resistances experienced during the first 1980s generation of AVM. The first service-wide implementation of RTPI in Dublin was for a new tram service in 2004, Luas, run on a franchise basis by Veolia, and before the bus implementation was finalised. It shares road-space with private vehicles for which it gets priority when approaching junctions. Sensors are placed at intervals of 100m or more which gather information from transponders on trams and send it to a central server and operations room.

As evident in the timeline below (Figure 1), Dublin Bus had accumulated or retained experience with RTPI through successive trials by the time of its definitive roll-out of AVL, in partnership with a specialist German transport technology company, INIT. A Dublin Transportation Office had been created in 1995 to put into place a transport strategy for the city region under the remit of the Department of Transport. It established a committee to oversee the implementation of RTPI in 2001, and contracted Atkins consultants in 2002 to create a general strategy. The report, published in 2006, noted the inconsistencies of information provision between services and the absence of an overarching mechanism to ensure holistic planning of both physical and informational infrastructure. It strongly recommended the creation of a specific public transport information office 'with responsibility for collecting data, publishing information and setting standards', 'the development and marketing of a public transport ''brand'' common to all modes and operators that the public can identify with, trust and rely upon', and 'the development of a set of agreements and processes governing agency participation' (Atkins, 2006: iv). The mooted idea of a land use and transport authority materialised in the form of the NTA, which fulfils the remit suggested in the report at the national scale. It also incorporated the Dublin Transportation Office and its data-modelling team (interview DSC27, NTA).

As the NTA was not created until 2008, 'it was agreed at the time [that the implementation of RTPI was initiated] that Dublin City Council was a better vehicle to actually procure the RTPI and subsequently that project and contract moved over to the NTA' (interview DSC27, NTA). The NTA have adopted the infrastructure created by Dublin City Council and added further measures to ensure its resilience, while also extending RTPI and the Leap smart travel card to other cities and towns in Ireland.

The AVL data from services are fed back into a central RTPI server, and subsequently into roadside display panels via cellular communication or GSM, with data hosted in two data centres in the Dublin Docklands with various failsafe mechanisms. This informs two official NTA apps (RealTime Ireland and Journey Planner) that cover all services, as well as several operator apps (Iarnród Éireann and Luas). RTPI data are available via a public API to researchers and commercial developers under a CCBY 4.0 license. Programmers can write their own queries in XML (extended mark-up language) or JSON (JavaScript object notation) to pull down specific information, which can then be pushed into custom-made displays for specific purposes.

These developments have given rise to two forms of data expansionism: that within and between the core operators themselves and largely based on AVL, including scheduling, data analytics and traffic prioritisation; and that permitted by the API, for research and commercial usage. The next section details how RTPI standards are managed between operators and the regulator in order to support this two-tier ecosystem, while the final empirical section considers how AVL and RTPI are supporting the operational consolidation of data-driven functionalities.

Negotiating and maintaining data standards

It was the responsibility of Dublin City Council, and then the NTA, to create consistency of information provision, ensure infrastructure compatibility, and develop a common brand with its accompanying clear aesthetic. This involved the redesign of bus-stops, the rollout of RTPI displays for multiple operators, the creation of an efficient and reliable back-end and telecommunications system, and the policing of information systems to ensure interoperability between transport providers. This required a co-constitutive development of procedures encompassing both human operators and software, where the latter is understood as partial or fully automated procedural systems.

Anomalies and breakdowns make infrastructure visible (Star and Ruhleder, 2001), and their resolution allows us to follow how standards are negotiated and stabilised. Such is the case with 'ghost buses', where

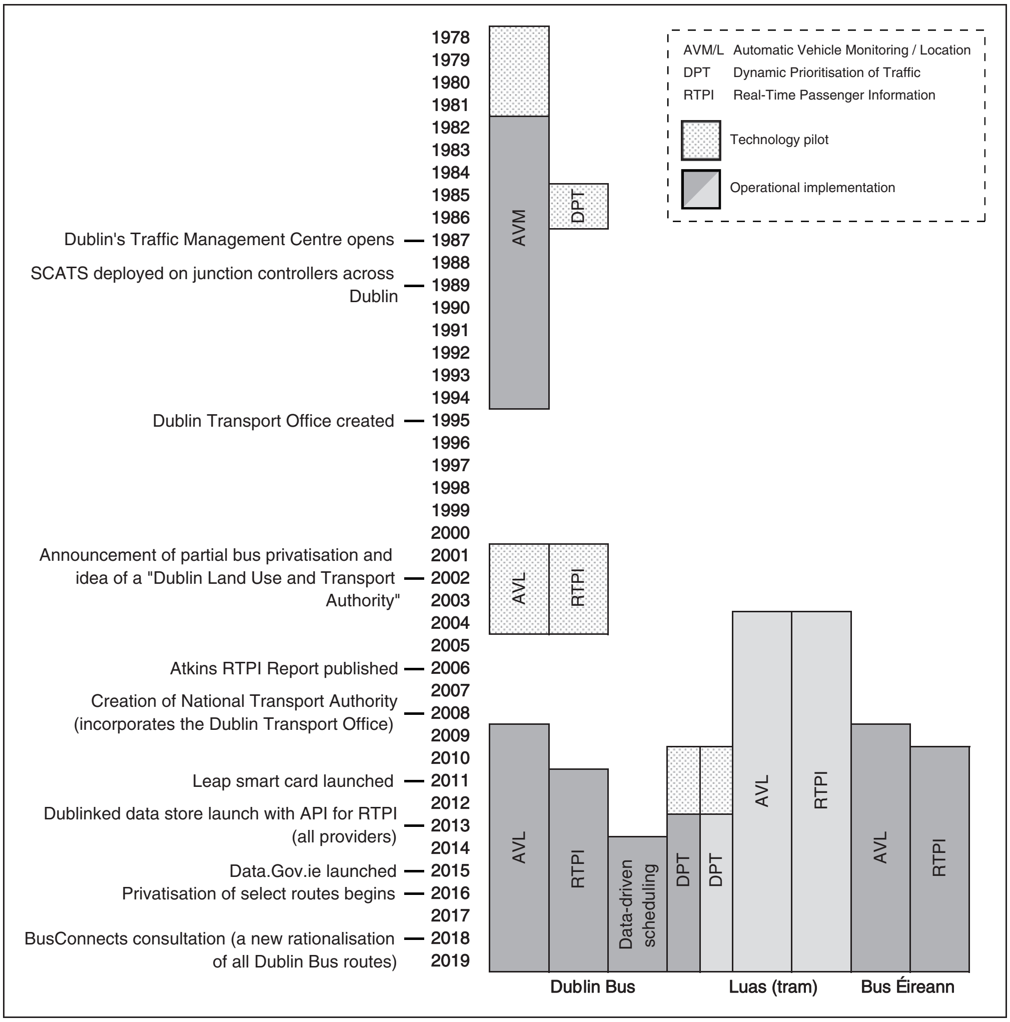


Figure 1. A timeline of RTPI-related transport technology deployments in Dublin for bus and tram services, with further events on the left. Iarnród Éireann is not included, which relies on its legacy signalling systems in addition to AVL.

services shown in RTPI fail to materialise. This anomaly provides insight into the various orders of issues encountered and their relationship to organisational change as the NTA exercised its authority over transport operators. There are many reasons for ghost buses, one of which was due to communication failures between operators as discussed below and resolved by early 2016. Its elimination and that of other issues that affected accuracy involved both repairing bugs and policing adherence to standards. For 2017, it was reported that 97.5% of RTPI information was correct with reference to arriving within 1 minute of the 'Due' prediction (Bus Átha Cliath, 2017), up from 92% in 2012 (Worrall, 2012) and 89% when initially launched.

In agreement with language policies for State bodies, RTPI information, including destination information, on trains, buses, and trams, is displayed and announced in both English and Irish. The original versions of many Irish place names now share equal space with their anglicised counterparts, e.g. *Cluain Saileach* (a meadow of willows) and 'Clonsilla'. The ghost bus anomaly resulted from a series of actions starting with Dublin Bus curtailing a bus and redirecting it before meeting its scheduled destination. Of the two information standards used by transport organisations in Ireland, VDV452 is used for scheduled timetables and SIRI for real-time feeds.[2] Dublin Bus changed a 19-character field in SIRI called 'DestinationRef' in a way that partners did not expect, as the employee cited below explains, because it did not have enough space for both languages, necessitating the use of a further field called 'JourneyNote':

> For example, we send the journey ref [which] would say, 'Okay, this journey is doing this destination and now that destination has changed'. So say, for example, a trip is going from A to B and we decide, okay, it is not going to B anymore, we'll bring it back somewhere.

Figure 2. Dublin bus live monitoring of schedule adherence in their control room.

> Now it can't handle that at the moment, the new destination, it still tells passengers you are going all the way to B. They are working on it and they are almost there. We had the same problem with the street signs for quite some time but they worked on that and they were fixed [by their external contractor] and that got resolved, that got changed. So now for example if you are going to Maynooth on a 66 [bus] for example, and for whatever reason, operational or whatever, they decide, okay this bus can't go the full journey, it can only go as far as Leixlip. We can now tell on the street signs, we can tell on our app on our website that to the passengers straight away, it could be on the bus itself, on the displays, they can be told this bus is now finishing in Leixlip. Whereas on the NTA's app they are still telling you 'Maynooth'. So it will be fixed quite soon, they have been working on it for a number of months now so it should be fixed soon. But on the street signs and our web that is reflected correctly. But that is kind of a bug thing, that isn't anything to do with the standards really, that is more a bug on their internal system than anything. It is no limitation of SIRI or the VDV that has caused that. (Interview RTPI01, Dublin Bus)

Dublin Bus considered that their usage was consistent with SIRI and that the issue reflected an external contractor's inability to read this information. In contrast, the NTA interviewee cited below indicated 'JourneyNote' as being more for minor notes like being 'guide dog-friendly'. This alleged non-standard use broke the information feed downstream, even if the change suited the challenge of frequently curtailing buses due to the large but necessary interruptions caused by Luas cross city.

A common solution for multiple language provision is to use a generic code in a look-up table with corresponding real-world names in various languages in adjacent columns:

> Yes, and largely that works fine if you don't go putting stuff in places where it shouldn't be, like journey notes shouldn't be in there. So the system trying to understand that doesn't see it and therefore these, I suppose important things, because curtailments and cancellations are really what you need to know about in real-time. You need to know if your bus is not going to the destination or if it has been cancelled. So they are things we can fix but they are expensive because what you will find is that an incumbent will see that as an opportunity to, let's say, know that nobody is going to compete with their price and therefore I would say give in ridiculously high prices to do fairly simple stuff. [. . .] And so really our experience is to kick back and say, 'Sorry, guys', right at the very start, stick to the specifications. [. . .] So over the years the system has been in place and more people want to build ancillary systems or reuse the information, the more we have learned that right at the start you need to be the policeman. (Interview SD13, NTA)

The central regulator talks of their role of policing common standards to ensure fair competition between providers through shared rather than privileged information. It has weekly meetings with RTPI representatives from the main transport providers during which anomalies and their potential solutions are discussed. While assertive about its policing role, a solution was negotiated that allowed Dublin Bus a degree of flexibility in their implementation of SIRI, creating custom code to recognise Dublin Bus curtailments. The changes were visible first on the street signs, and reflected later on the NTA smartphone app.[3]

This second-order issue on adherence to common standards reflects a transition from putting in place the basic infrastructure towards consolidating RTPI in a new suite of procedures and processes, yet also interacts with higher third-order reconfigurations of power relations as the regulator imposes a new performance-driven ethos on operators. As the transport franchise-holder, the NTA has the power to outsource routes to other suppliers, oversees performance management for all contracted transport providers, and is therefore at the heart of negotiations on privatisation and standardisation. During the time of fieldwork, there were industrial disputes involving Luas tram drivers, and protracted discussions on the costs and benefits of partial privatisation of national services targeting Dublin Bus and Bus Éireann. Over the course of 2016 to 2018, 24 routes from Dublin Bus and 6 routes from Bus Éireann were listed for privatisation (*Tallaght News*, 2017) with UK firm Go-Ahead winning both competitions. The NTA manages routes, schedules, fares, vehicles, quality control and RTPI for new providers. In this manner, it hopes that like with Transport for London, tenders for routes are largely inconsequential for passengers through a model of competition *for* the market rather than *in* the market (Preston and Almutairi, 2013).

Operational consolidation here was the prerogative of an increasingly assertive state actor, the NTA, which addressed administrative fractures between State companies through control over both markets and data. This facet of data ratcheting comprises data-driven organisational control over operators and requires collective adherence to data policies. At the intra-organisational scale, within Dublin Bus and Bus Éireann, real-time data on the location of vehicles and drivers also enabled data-driven organisational control, in part based on new data-driven functionalities. This ratcheting provides momentum to the cycle of expansionism, experimentalism and consolidation as new datasets are created and further use-cases found.

Data-driven procedural change and data ratcheting

AVM and RTPI technologies in Dublin have given rise to a cascade of data-driven organisational changes. In many instances, these are local learnings of established international best-practices, yet also include instances of innovation such as the many small data-driven changes to routes to reach industry-standard RTPI accuracy. Dublin Bus's schedule management software is reliant on AVL and creates a more constant and intimate connection between drivers and controllers. The latter are housed in a control centre and seated at cubicles equipped with several graphical user interfaces for monitoring buses in real-time.

Much the same as the recording devices affixed to chain retail workers, these technologies constrain employee behaviour in order to provide an experience to customers that is invariable and therefore negotiable with less cognitive effort. Dublin Bus is a public company and many of the staff in the control room are ex-drivers themselves, their exchanges with drivers replete with banter and laughter yet nevertheless effecting a culture change within the organisation:

> [B]efore the AVL system, the only way of knowing where a bus was, unless there was a guy out on the street watching the buses coming in and the inspector on the road. Or else it was the control up above calling the driver and saying, 'Where are you?' And now it is a case of . . . It was funny at first seeing it happen because they'd be saying I am in such and such a street. And the controller would say, 'No you are not. I can see within 20 seconds of where you are'. That took a bit of getting used to and now the drivers are at the point where they are embracing it very much so. It took a while for them to trust it. [. . .] But yeah, they have taken it on board now and they do realise all the old ways of operating have to change. (Interview RTPI01, Dublin Bus)

The interviewee above notes the role of AVL in supporting measures to improve quality of service. They are enforced not only by control room operators, on whose terminal screens buses are highlighted in red for being behind or ahead of schedule (see Figure 2), but also on vehicles themselves through devices to notify drivers if they are running ahead of their scheduled stops via a beep and a red dashboard light (only when the vehicle is stationary). Together with the control room and its real-time displays, data analytics, and a management team with key performance indicators to meet, these technologies all coerce human behaviour towards minimising error and irregularities. A realisation of the potential of data to drive internal reform has led to further local innovations, including using data analytics packages to recalibrate schedules (timetables) based on statistically examined journey times, and developing new internal performance metrics for management. These measures further improve RTPI accuracy, and evidence how data ratcheting reshapes internal organisational control as engineers experiment with, and operationalise, new functionalities.

In addition to schedule and route alterations, further interventions in the interest of efficiency can be either physical, such as changes to the road layout, or digital, through alterations to the junction signalisation patterns of traffic lights. The greater part of Dublin's road network is managed with industry-standard software (SCATS) that alters junctions dynamically in response to traffic by swapping between pre-programmed plans. Traffic prioritisation has now been rolled across the SCATS network for public transport by using RTPI data in SIRI format to estimate proximity (O'Donnell et al., 2018). This belated but eventually successful deployment of traffic prioritisation (see Figure 1) relates to third-order issues surrounding the dominant trend of transitioning to a low-carbon future by progressively revoking privileges once handed out willingly to the private motorist and increasing investment into public transport.

The schedule and fleet management software, the data analytics functions for route optimisation, and automated bus prioritisation form part of an established practice of ratcheting the recognised power of data to redefine procedures, and by extension, urban flows and spatial relations. Data-driven functionalities become perceived as integral during this phase of efficiency and rationalisation. A Dublin Bus technician notes that data are 'growing more and more into decision-making around the company' and adds:

> So it is a complete [change of] mind-set, everything has changed. Once you trust in the data better decisions can be made [. . .].
> It was unbelievable straight away, things like when a door opened, you could see straight away, just information that you just couldn't have before so straight away you were just giving the power to the managers to see exactly what was going on, the routes, and ultimately that is getting back to the customers. (Interview RTPI04, Dublin Bus)

This localised data experimentalism and consolidation occur alongside larger-scale operational changes as centralised authorities coordinate transport services and data, creating new products and services to consolidate those initially created by individual transport providers. This includes the present BusConnects programme, initiated in 2017 and led by the NTA, for the rationalisation of all bus services in Dublin. Two NTA participants below recognise the potential of data for better strategic planning:

> So I think once you have the kit with proper management and working between different companies you can make these improvements to raise the reliability and the perception of reliability to people. So I see expanding the system but also harvesting it to offer better products. (Interview SD13, NTA)
> Yes, we use the data for a number of purposes. The initial objective was to provide more information to the consumers to actually make transport more accessible; that is the real value we saw from the information that we were collecting and providing. But certainly we have realised now that we have a huge wealth of information at our disposal in terms of where buses are at, the best journey patterns being used, and we can analyse that now and start using that to make the planning decisions more effective going forward. (Interview DSC27, NTA)

The expansion of metrics and accountability reinforces an 'audit culture' (Strathern, 2000), dependent on technologies that ostensibly serve passenger service provision but which also expand into technologies of control and regulation. Data ratcheting represents a quasi-Lamarckian evolutionary progression as organisations create ever more sophisticated secondary products from high-quality data, recognising its power and curating it with discretion according to socio-cultural specificities and legal frameworks. On the one hand, this may involve preventing such data from being misused for criminal purposes, while on the other, it could be used to maintain competitive advantage or increase control over consumers or citizens. The empirical data presented here account for largely internal and inter-organisational data-driven innovations yet may eventually lead to broader data assemblages beyond the transport sector, aligning with broader commercial or political interests.

Conclusion

In their future roadmap of ubiquitous computing, Dourish and Bell (2011) highlight the emergent nature of technologies, which initially developed separately, coalesce in what are seen retrospectively as unitary systems such as the smart home or the city dashboard (Kitchin et al., 2016). Technological changes are also essentially organisational and procedural, liberating the human mind from repetitive drudgery, and facilitating new forms of behaviour between people and their interaction with the environment. Tracing their evolutionary development reveals the multiple path dependencies and reverse salients (Hughes, 1993) that characterise their real-world implementation, and furthermore illustrates the value of blending the historicism of infrastructure studies with the study of data assemblages.

The rollout of RTPI in Dublin is interrelated with broader transport technologies including automated vehicle location systems and traffic control systems, and evidences how its local implementation is part of an arc of technological changes in ITS. It initiates in the 1970s with advanced trials that were inconsistently resourced by the State until the context itself had shifted. The national economy had moved towards higher-value technologies and services with a more informed and demanding workforce, necessitating the provision of seamless multimodal transport, integrated ticketing, and RTPI.

The deployment of RTPI allowed individual operators such as Dublin Bus to experiment with and drive their own internal procedural reforms, with instances of both localised learnings and more genuine innovations as local actors participated in European and international networks for developing and disseminating best practices. These activities are described here as instances of data ratcheting, as new data-driven functionalities are implemented iteratively as they are discovered in a local context, oriented towards a more rationalised data-driven culture. Glitches such as 'ghost buses' provide a window into the politics of data, as the transport sector modernised and the NTA consolidated its mandate to improve passenger experience by policing standards and controlling providers. This hierarchy of procedural authority, from regulator to operator to driver, has been strengthened through the creation of governmental bodies that ensure seamless exchange of data and the operational consolidation of data-driven procedures. In addition to favouring a rationalised audit culture (Strathern, 2000), this will likely assist in future data-driven functionalities that integrate with other sectors and services.

Urban datafication in transport has been characterised here as consisting of the three translations of data expansionism, data experimentalism and operational consolidation. It is tentatively argued, pending further comparative reviews of in-depth local studies, that there will be an increasing tendency to curate and streamline data as the potentialities of Big Data shared across multiple domains are appreciated and realised. Transport operators, particularly those serving massive urban populations, can use tracking technologies for efficiency, commerce and security. Transport for London, for example, while maintaining their open data portal and subject to comprehensive UK and EU data protection and privacy regulations, are investing in large-scale tracking technologies to inform their services (McMullan, 2018; Sweeney, 2018). Transport data may be combined with social networks and public services profiles to perform new forms of citizenship as smartphone-delivered personalised notifications and services become normalised. Further research could enquire as to the multiple data streams that are becoming operational, their epistemologies, and their primary benefactors. For instance in China, transport apps for citizens are being tied to unique identifiers that may restrict or incentivise transport options according to their government-measured 'social credits' (Carney, 2018). Such developments may similarly benefit from the comprehensive approach of data assemblages that contextualise data-driven technologies in their socio-cultural contexts, as well as from the extended time-scales of infrastructure development and attention to glitches and breakdowns.

Acknowledgements

I am indebted to Professor Rob Kitchin for his very helpful feedback and input into this article and research, which formed part of a larger research project on the smart city for which he was principle investigator. My thanks also to Dr Sheila Convery for her comments on the article, and to the editorial team and three anonymous reviewers for their comments.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research for this article was provided by a European Research Council Advanced Investigator Award, 'The Programmable City' (ERC-2012-AdG-323636).

ORCID iD

Liam Heaphy https://orcid.org/0000-0003-1452-4369

Notes

1. A clear instance of technology-dependent automation is driverless underground trains in cities like Paris (Lines 1 and 14), where the speed and intervals between trains are tightly controlled through a central system and arrays of sensors and machinery. In this case, they become code/spaces (Kitchin and Dodge, 2011) in which software and space are mutually constitutive.
2. Service Interface for Real-time Information (SIRI) is a European standard in the form of an extended mark-up language (XML) for distributing real-time transport information developed by a partnership of European transport bodies (see: http://www.transmodbku.kiljikiljik;/liilulel-cen.eu/standards/siri/, accessed 13 November 2018). VDV stands for the German *Verband Deutscher Verkehrsunternehmen* (Association of German Transport Companies), who have created three timetabling standards. VDV452 is the main planned timetable, while VDV453 is stop-centric, providing real-time updates per stop. VDV454 is route-centric, reflecting scheduled times on a given route.
3. According to the SIRI documentation (https://www.vdv.de/siri.aspx) and its implementation elsewhere, it would seem for the Estimated Timetable function in SIRI, JourneyNote is by design indicated primarily, but by no means exclusively, for additional information such as being wheelchair friendly rather than destination information, which corresponds instead to the DestinationRef and DestinationName fields.

References

Atkins (2006) *A Strategy for Integrated Public Transport Information, including Real Time, in the Greater Dublin Area*. Dublin, Ireland. Available at: www.nationaltransport.ie/downloads/archive/atkins_report_2006.pdf (accessed 7 August 2018).

Bus Átha Cliath (2017) *Bus Átha Cliath Annual Report and Financial Statements 2017*. Available at: www.dublinbus.ie/About-Us/Reports/Annual-Reports/ (accessed 11 April 2019).

Cardullo P and Kitchin R (2018) Being a 'citizen' in the smart city: Up and down the scaffold of smart citizen participation in Dublin, Ireland. *GeoJournal* 84: 1–13.

Carney M (2018) Leave no dark corner. *ABC News*, 18 September. Available at: www.abc.net.au/news/2018-09-18/china-social-credit-a-model-citizen-in-a-digital-dictatorship/10200278 (accessed 21 September 2018).

Caulfield B and O'Mahony M (2003) Real time passenger information: The benefits and costs. In: *The association for European Transport Conference*, Strasbourg, 8 October 2003. Available at: http://www.tara.tcd.ie/bitstream/handle/2262/79614/Caulfield O'Mahony%2C Real time passenger information costs and benefits.pdf (accessed 29 July 2019).

Caulfield B and O'Mahony M (2004) The provision of on street passenger information via real time passenger information: A case study of Dublin. In: *12th IEE international conference on road transport information and control, 2004*. RTIC 2004. April 2004. DOI: 10.1049/cp:20040001

Caulfield B and O'Mahony M (2009) A stated preference analysis of real-time public transit stop information. *Journal of Public Transportation* 12(3): 1–20.

Coletta C, Heaphy L and Kitchin R (2018a) From the accidental to articulated smart city: The creation and work of 'Smart Dublin'. *European Urban and Regional Studies* 1–16. Available at: https://doi.org/10.1177/0969776418785214 (accessed 29 July 2019).

Coletta C, Heaphy L, Perng S-Y, et al. (2018b) Data-driven cities? Digital urbanism and its proxies: Introduction. *TECNOSCIENZA: Italian Journal of Science & Technology Studies* 8(2): 5–18.

Dourish P and Bell G (2011) *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. Cambridge, MA: MIT Press.

Edwards PN (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Foth M, Brynskov M and Ojala T (eds) (2015) *Citizen's Right to the Digital City: Urban Interfaces, Activism, and Placemaking*. Singapore: Springer.

Giffinger R and Pichler-Milanović N (2007) *Smart Cities: Ranking of European Medium-sized Cities*. October. Vienna: Vienna University of Technology.

Gomez-Ibanez J and Meyer JR (2011) *Going Private: The International Experience with Transport Privatization*. Washington, DC: Brookings Institution Press.

Gray J, Gerlitz C and Bounegru L (2018) Data infrastructure literacy. *Big Data & Society* 5(2): 1–13.

Hughes TP (1993) *Networks of Power: Electrification in Western Society, 1880–1930*. Baltimore, MD: The Johns Hopkins University Press.

Kitchin R (2014a) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 1–12.

Kitchin R (2014b) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: Sage.

Kitchin R, Coletta C and Heaphy L (2018) Actually existing smart Dublin: Exploring smart city development in history and context. In: Karvonen A, Cugurullo F and Caprotti F (eds) *Inside Smart Cities: Place, Politics and Urban Innovation*. London: Routledge, pp. 83–101.

Kitchin R and Dodge M (2011) *Code/Space: Software and Everyday Life*. Cambridge, MA: MIT Press.

Kitchin R and Lauriault TP (2014) *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work*. ID 2474112, SSRN Scholarly Paper, 30 July. Rochester, NY: Social Science Research Network. Available at: http://papers.ssrn.com/abstract=2474112 (accessed 12 December 2018).

Kitchin R, Maalsen S and McArdle G (2016) The praxis and politics of building urban dashboards. *Geoforum* 77: 93–101.

McMullan T (2018) Transport for London to start automatically tracking bus passenger numbers. *Alphr*, 7 June. Available at: www.alphr.com/go/1009545 (accessed 12 November 2018).

Meinel C (2004) Molecules and croquet balls. In: Chadarevian SD and Hopwood N (eds) *Models: The Third Dimension of Science*. Illustrated edition. Stanford, CA: Stanford University Press, pp. 242–275.

NTA (2018) *Annual Report & Financial Statements 2017*. Rialtas na hÉireann. Available at: www.nationaltransport.ie/wp-content/uploads/2018/12/NTA_AR_-_FS_2017_ENGLISH_FINAL.pdf (accessed 11 April 2019).

O'Donnell M, Bolger N and Calvey J (2018) Data driven decisions for centralised bus priority in Dublin City. In: *JCT Traffic Signals Symposium & Exhibition*, Nottingham, September 2018, p. 15. Available at: www.jctconsultancy.co.uk/Symposium/Symposium2018/symposium2018.php (accessed 10 December 2018).

Perng S-Y and Kitchin R (2016) Solutions and frictions in civic hacking: Collaboratively designing and building wait time predictions for an immigration office. *Social & Cultural Geography* 19: 1–20.

Preston J and Almutairi T (2013) Evaluating the long term impacts of transport policy: An initial assessment of bus deregulation. *Research in Transportation Economics* 39(1): 208–214.

Reed DJ (2018) Discovering the data-driven city. Breakdown and literacy in the installation of the Elm Sensor Network. *TECNOSCIENZA: Italian Journal of Science & Technology Studies* 8(2): 81–104.

Reed DJ and Webster A (2010) Architectures of motility: ICT systems, transport and planning for complex urban spaces. In: Nunes Silva C (ed.) *Handbook of Research on E-Planning: ICTs for Urban Development and Monitoring*. Hershey, PA: Information Science Reference, pp. 365–387.

Roth SH (1977) History of automatic vehicle monitoring (AVM). *IEEE Transactions on Vehicular Technology* 26(1): 2–6.
Shapiro A (2018) The urban stack. A topology for urban data infrastructures. *TECNOSCIENZA: Italian Journal of Science & Technology Studies* 8(2): 61–80.
Shelton T (2017) The urban geographical imagination in the age of Big Data. *Big Data & Society* 4(1): 1–14.
Shelton T, Zook M and Wiig A (2015) The 'actually existing smart city'. *Cambridge Journal of Regions, Economy and Society* 8(1): 13–25.
Sørensen CH and Gudmundsson H (2010) The impact of governance modes on sustainable transport – The case of bus transport in Greater Manchester, UK. *World Review of Intermodal Transportation Research* 3(1): 8–25.
Star SL (1999) The ethnography of Infrastructure. *American Behavioral Scientist* 43(3): 377–391.
Star SL and Bowker GC (2006) How to infrastructure. In: Lievrouw LA and Livingstone S (eds) *Handbook of New Media: Social Shaping and Social Consequences of ICTs*. Updated student edition. Thousand Oaks, CA: Sage Publications, pp. 230–245.
Star SL and Ruhleder K (2001) Steps toward an ecology of infrastructure: Design and access for large information spaces. In: Yates J and van Maanen J (eds) *Information Technology and Organizational Transformation: History, Rhetoric, and Practice*. Thousand Oaks, CA: Sage Publications, pp. 305–346.
Strathern M (ed.) (2000) *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*. London: Routledge.
Sweeney R (2018) Opening data fully to improve London's transport network. *Intelligent Transport*, 26 February. Available at: www.intelligenttransport.com/transport-articles/65404/opening-data-full-capacity-transport/ (accessed 12 November 2018).
Symes DJ (1980) Automatic vehicle monitoring: A tool for vehicle fleet operations. *IEEE Transactions on Vehicular Technology* 29(2): 235–237.
Tallaght News (2017) Sean Crowe has described the privatisation of 24 Dublin Bus routes as worrying, 18 August. Available at: www.newsgroup.ie/sean-crowe-has-described-the-privatisation-of-24-dublin-bus-routes-as-worrying/ (accessed 9 April 2018).
Watkins KE, Ferris B, Borning A, et al. (2011) Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice* 45(8): 839–848.
Wiig A (2019) Incentivized urbanization in Philadelphia: The local politics of globalized zones. *Journal of Urban Technology* 1–19. Epub ahead of print 2019. Available at: https://doi.org/10.1080/10630732.2019.1573628 (accessed 29 July 2019).
Williams B (2008) *Intelligent Transport Systems Standards*. Boston, MA: Artech House.
World Bank (2011) Intelligent Transport Systems – Dublin, Ireland. Available at: www.ssatp.org/sites/ssatp/files/publications/Toolkits/ITS%20Toolkit%20content/case-studies/dublin-ireland.html (accessed 13 December 2018).
Worrall JJ (2012) For travel innovators, timing is everything. *The Irish Times*, 5 July. Available at: www.irishtimes.com/business/technology/for-travel-innovators-timing-is-everything-1.529501 (accessed 13 December 2018).
Yates J (1993) *Control through Communication: The Rise of System in American Management*. Johns Hopkins paperbacks ed. Studies in industry and society 6. Baltimore, MD: Johns Hopkins University Press.

Editorial

BIG DATA & SOCIETY

Big Data & Society
July–December 2019: 1–6
© The Author(s) 2019
DOI: 10.1177/2053951719867349
journals.sagepub.com/home/bds
SAGE

How should we theorize algorithms? Five ideal types in analyzing algorithmic normativities

Francis Lee[1] and **Lotta Björklund Larsen**[2]

Abstract

The power of algorithms has become a familiar topic in society, media, and the social sciences. It is increasingly common to argue that, for instance, algorithms automate inequality, that they are biased black boxes that reproduce racism, or that they control our money and information. Implicit in many of these discussions is that algorithms are permeated with normativities, and that these normativities shape society. The aim of this editorial is double: First, it contributes to a more nuanced discussion about algorithms by discussing how we, as social scientists, think about algorithms in relation to five theoretical ideal types. For instance, what does it mean to go under the hood of the algorithm and what does it mean to stay above it? Second, it introduces the contributions to this special theme by situating them in relation to these five ideal types. By doing this, the editorial aims to contribute to an increased analytical awareness of how algorithms are theorized in society and culture. The articles in the special theme deal with algorithms in different settings, ranging from farming, schools, and self-tracking to AIDS, nuclear power plants, and surveillance. The contributions thus explore, both theoretically and empirically, different settings where algorithms are intertwined with normativities.

Keywords

algorithms, theory, normativities, black boxing, infrastructures, actor-network theory

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

The omnipresence of algorithms

Algorithms are making an ever-increasing impact on our world. In the name of efficiency, objectivity, or sheer wonderment algorithms are becoming increasingly intertwined with society and culture. They seem pervasive in today's society, materializing here, there, and everywhere. In some situations, algorithms are valued, or even treasured, in others they lead to anger and mistrust, sometimes they even seem threatening or dangerous.

In the wake of this algorithmization of the world, social scientists have taken an increasing interest in how algorithms become intertwined with society and culture. The list of interventions and perspectives seems endless.[1] Some researchers claim that algorithms control money and information (Pasquale, 2015) or shape our romantic endeavors (Roscoe and Chillas, 2015). Others highlight the inscrutability of algorithms and work to understand the effects of their opacity (Burrell, 2016; Diakopoulos, 2016; Fourcade and Healy, 2017; Pasquale, 2015). Still others argue that algorithms automate inequality (Eubanks, 2017; Noble, 2018; O'Neil, 2016), and reproduce existing social structures and biases (Angwin et al., 2016; Kirkpatrick, 2016; Sandvig et al., 2016). In line with

[1]Uppsala Universitet, Uppsala, Sweden
[2]Tax Administration Research Centre, University of Exeter Business School

Corresponding author:
Francis Lee, Uppsala Universitet, Uppsala 75105, Sweden.
Email: francis@francislee.org

this, many researchers have started asking questions about algorithmic decision-making (Zarsky, 2015), accountability (Diakopoulos, 2016), or ethics (Kraemer et al., 2010; Neyland, 2018). Implicitly, or sometimes very explicitly, many of these observe that algorithms are intertwined with different normativities and that these normativities come to shape our world.

In our view, however, there is a need for a meta-discussion about how normativities become intertwined with algorithms. That is, there is today a bounty of approaches, as evident from above, and a growing body of analytical perspectives on algorithms. However, there are few, if any, meta-analytical discussions that attempt to deal with the strengths and weaknesses of different theoretical and analytical approaches. Consequently, this special theme seeks to provoke a meta-reflection on how social and cultural researchers come to theorize, analyze, and understand algorithmic normativities.

Five ideal types in analyzing algorithms

In essence, what we are asking is what alternatives there are to emphasizing opacity, black-boxing, and a seductive language of uncovering 'biased', 'racist' or 'sexist' algorithms? What other ways of approaching algorithms are there apart from going under the 'opaque' hood to understand the real politics of the 'black boxed' algorithms? To stimulate this meta-reflection, and help think these issues through, we want to draw playfully on the metaphor of the engine hood. We ask what positions, other than 'going under the hood' to uncover the hidden normativities of the algorithm, are there?

The goal of this exercise is to reflect on and problematize current ways of analyzing algorithms, and to situate the articles in this special theme in relation to a meta-reflection about how we analyze algorithmic normativities.[2] In doing this, we outline five ideal-typical accounts of algorithms in society and situate them in relation to classical analytical positions in Science and Technology Studies (STS), where debates about the social analysis of technology abound. In outlining these analytical ideal types, we also summarize and situate each article in this special theme in relation to these ideal types.

We are aware that this strategy risks foregrounding some perspectives while backgrounding others. There is also a risk that we become too harsh in our ideal typing by omitting and downplaying overlaps and similarities between different perspectives. So, bear with us if we do violence to more nuanced and multifaceted perspectives when we pull some things apart and push some other things together.

Under the hood: The politics of algorithms

Let us start by going under the hood. A number of researchers maintain that we must analyze algorithms themselves and go under the hood to understand their inherent politics (cf. Ruppert et al., 2017). Several social scientists interested in algorithms or Big Data gather in this camp. For instance, philosophers sometimes like to debate the algorithmic ethics of self-driving cars (Nyholm and Smids, 2016), or racial bias in facial recognition (Chander, 2017), while other analysts have dealt with the algorithmic biases of criminal risk prediction (Angwin et al., 2016).

In this analytical ideal type, the logic of the algorithm appears like a *deus ex machina* impinging on society's material politics. This analytical ideal type draws on similar logics to Langdon Winner's (1980) classic article that deals with the politics of technological artifacts. In his analysis, it is the functioning of the technological system that is in focus, and Winner invites us to see artifacts as materialized laws that redefine how we can act for generations to come.

Algorithms are also productively and provocatively understood in this way. For instance, in Christopher Miles contribution to this special theme, he fruitfully illustrates how algorithms become intertwined with specific normativities in American farming. Miles shows that although new digital practices are introduced, existing socio-economic normativities are often preserved or even amplified. Thus, the supposedly radical changes in farming practices, suggested by precision agriculture algorithms, only affect certain parts of farming but also seem to thwart other imagined futures.

In this type of analysis of algorithms, just as in the case with Winner's bridges, the politics, effects, and normativities that are designed into algorithms become foregrounded. A crucial task for the researcher thus becomes to analyze invisible algorithmic normativities to understand how they are imbued with power and politics. However, with this type of analysis, might we risk losing sight of the practices, negotiations, and human action that algorithms always are intertwined with? Might we become so seduced by the algorithms, that we forget the many social practices that surround them?

Working above the hood: Algorithms in practice

What would happen if we instead stay above the hood and never get our hands dirty with the nuts and bolts of the inscrutable algorithms? Here, on the other side of our constructed spectrum of ideal types, we could place ethnomethodological analyses of the achievement of social order. In this analytical position, algorithms would emerge as 'contingent upshot of practices, rather than [as] a bedrock reality' (Woolgar and Lezaun, 2013: 326).

In terms of classic studies of materiality in social interaction, Charles Goodwin's (2000) analysis of the classification of dirt might serve as emblematic. In his article, Goodwin analyzes the interaction, talk, pointing, disagreement, that happens when archaeologists attempt to classify dirt color using a so-called Munsell-chart. Goodwin shows how the human practices of interpreting and debating is crucial to understanding how the meaning of material artifacts is decided.

In relation to algorithms, Malte Ziewitz' (2017) account of an 'algorithmic walk' also humorously points us toward the negotiations that surround most algorithms. He highlights the constant work of interpreting, deciding, and debating about algorithms. In an auto-ethnographic experiment, Ziewitz takes a stroll with a friend in the streets of Oxford and makes it into an algorithmic walk. Before the walk, Ziewitz and his friend construct strict algorithmic rules for how to turn at intersections during the walk. But walking like an algorithm turns out to be a difficult task, and Ziewitz and his friend soon start debating the meaning of the algorithm. What direction does the algorithm want us to go at this intersection?

In this vein, Lotta Björklund Larsen and Farzana Dudwhala's contribution to this special theme also discusses how humans adapt to algorithmic outputs, an adaptation that includes normative ideas about how the outcomes of algorithms are interpreted as normal or abnormal. Their argument is that the output of algorithms trigger, or as they propose, *recalibrate* human responses. Sometimes accepting the algorithmic output, sometimes not, so that an understanding of a normal situation can be achieved.

In another article in this special theme, Patricia DeVries and Willem Schinkel also undertake an analysis of the practical politics of algorithms. In their article, they analyze how three artists construct anti-facial-recognition face masks to critique and resist facial recognition systems. They take these face masks as a starting point for exploring the social anxiety around surveillance and control. DeVries and Schinkel argue that there is a tendency, in these works, to defend a liberal modernist construction of the autonomous subject and the private self. The meaning of both the face masks and the surveillance algorithms are thus negotiated in practice.

This analytical stance elegantly sidesteps the discussion about which nefarious or biased politics are designed into the algorithm. The human negotiations drawing on contexts, materialities, or even face masks, become foregrounded. Opening the opaque and black boxed algorithm to decode its inscrutable politics becomes almost irrelevant; it is the interpretation in practice that is in focus. However, do we then, by working above the hood, risk omitting what algorithms are actually constructed to do? And, we might ask, what then is the price of staying over the hood to focus on practice?

Hoods in relations: Non-human agency and black boxes

This brings us to the ideal type that approaches algorithms, and technology, through an analysis of non-human agency and relationality—perhaps a middle road between going under the hood and staying above it? This analytical ideal type focuses on the intertwining of human and non-human actors (cf. Callon and Law, 1995).[3] Here, for instance, Actor-Network Theory (ANT) in its various guises zooms in on the effects of how both non-human and human actors are intertwined (Latour, 1987).[4]

Relationalities are, for example, the focus of the article by Francis Lee, Jess Bier, Jeffrey Christensen, Lukas Engelmann, Claes-Fredrik Helgesson, and Robin Williams. The authors criticize the current focus on fairness, bias, and oppression in algorithm studies as a step toward objectivism. Instead, they propose to pay attention to *operations of folding* to highlight how algorithms fold a multitude of things, such as data, populations, simulations, or normalities. For instance, they show how the algorithmic calculation of the spread of an epidemic can produce particular populations or countries as being close to an epidemic, while others seem safely distant. The algorithmic production of different relations thus having potentially far-reaching consequences for both individuals and nations.

Similarly, in a comment on accountability and authorship, Elizabeth Reddy, Baki Cakici, and Andrea Ballestero highlight, through the algorithmic experiments of a comic book artist, how algorithms can do much of the work of assembling detective stories. Yet, accountability is still organized, normatively and legally, around human authorship and human agency. So, although algorithms might produce 'the work,' ideas about authorship and accountability are still organized around human subjectivity and agency. They observe that there seems to be a tendency to ascribe agency to humans over machines (cf. Callon and Law, 1995).

In this ideal type, by focusing on the practices of relational ordering, we can come to understand the complex mixing of agencies and accountabilities between algorithms and humans. Instead of seeing black boxed algorithms as a delimiter for a study, this position sees it as the starting point for inquiry.[5] Attempting to discern both how the algorithm functions and how it relates to human practice is the *modus operandi*. Perhaps one might describe this ideal type as combining 'going under the hood' with the practice oriented 'staying above the hood' ideal types. However, this relational perspective has also been criticized for being apolitical and blind to the power-struggles of weaker actors, as well as the political effects that algorithms could have on the world (cf. Galis and Lee, 2014; Star, 1991; Winner, 1993).

Lives around hoods: Torque, classification, and social worlds

Let us widen the lens even more: From detailed studies of interaction, negotiation, and relationality to an analysis of neighborhoods (da-dum-tsch). This analytical position takes an interest in infrastructures of classification and their interaction with human biographies. Here, the politics of infrastructures and classification become the focus. These types of analyses highlight how people's lives become 'torqued', or twisted out of shape, by classification systems.[6]

For instance, Bowker and Star (1999), in their classic work on infrastructures of classification, show how the socio-material race classification systems of the South African apartheid regime affected human lives in sometimes unpredictable ways. For instance, a light brown child born to dark brown parents could be forced to go to school in another district. But they also show how neighborhoods could come together to challenge the classification system—as the definition of race in apartheid South Africa was founded on social and legal negotiations.

In moving this ideal type to the algorithmic arena, a thought-provoking approach might be Philip Roscoe's (2015) study of a kidney transplant algorithm in the UK. He shows how the question 'Who is a worthy recipient of a kidney?' is answered in algorithmic form. But the algorithm—just as apartheid race classifications—is tied to the valuations of worthy recipients. Furthermore, just as neighbors could sometimes band together to challenge a color classification in South Africa, so can hospital staff today 'game' the algorithm for a 'worthy' recipient, while the algorithm is still used as an escape route from making heart-wrenching decisions on life and death.

Such negotiated processes of classification are also the center of attention in Helene Gad Ratner and Evelyn Ruppert's article in this special theme, which analyses the transformation of data for statistical purposes. In their text, they show how metadata and data cleaning as aesthetic practices have normative effects. Just as Bowker and Star (1999) dealt with the struggles of, for instance, medical, biological, or racist classifications, Ratner and Ruppert show how classification struggles happen through the practices of data cleaning. One instance they document is how absences and indeterminacies in data are resolved through both algorithmic and human interactions with the data. Importantly, these interactions determine what values the data can obtain. Thus, similar to how the apartheid regime performed the population of South Africa, Ratner and Ruppert bring analytical awareness to the normative and performative effects of classification work in relation to homeless and student populations and how they are enacted by infrastructures.

In this ideal type, the work of classification and the relations between human lives and classification systems becomes foregrounded. Here the politics of twisting lives out of shape becomes the focus. However, perhaps a risk is that we lose sight of the detailed interactions of how social order is maintained in practice? Or does a focus on the algorithms of classification risk leading us back, full circle, to seeing algorithmic systems as having inherent politics again?

The mobile mechanics: The power of analytical reflexivity and mobility

Finally, we wish to highlight how a meta-reflexive and meta-analytical attitude toward algorithms opens new avenues for inquiry. By being attentive to how social scientists relate to algorithms as well as to those who work with them, our inherent normativities and presumptions come to the fore.

In this special theme, two articles analyze such interventions. David Moats and Nick Seaver challenge our thinking about how computer scientists understand the work of social scientists. In the article, Moats and Seaver document their attempt to arrange an experiment with computer scientists to test ingrained boundaries: how can the quantitative tools of computer science be used for critical social analysis? As it turns out, the authors were instead confronted with their own normative assumptions. By sharing these insights, the authors provoke the reader's assumptions about the normative disparities inherent in different scientific disciplines.

Last, and in a similarly reflexive approach, Jeremy Grosman and Tyler Reigeluth propose a meta-framework for understanding algorithmic normativities. Discussing the notion of normativity from the point of view of various analytical positions, algorithmic systems are said to produce three kinds of normativities: technical, sociotechnical, and behavioral. The authors argue that algorithmic systems are inhabited by *normative tensions* and that a fruitful approach is to explore the tensions instead of the normativities themselves. Their argument is that this approach allows them to show which norms get more traction than others and perhaps even suggest why this is so.

Conclusion

A point of departure for this special theme was that algorithms are intertwined with normativities at every step of their existence; in their construction, implementation, as well as their use in practice. The articles explore theoretically and empirically different settings where humans and non-humans engage in practices that are intertwined with normative positions or have normative implications. The array of theoretical approaches—anxieties, pluralities, recalibrations, folds, aesthetics, accountability—that implicate algorithms force us to engage with the multiple normative orders that algorithms are entangled with.

In the articles, we get to see how algorithms are intertwined with, on the one hand, expectations of how things ought to be—normative expectations—and, on the other hand, how they enact the normal, the typical, as well as the abnormal and atypical. The articles thus scrutinize ideas of normativities in and around algorithms: how different normativities are enacted with algorithms, and how different normativities are handled when humans tinker with algorithms.

With this brief editorial, we hope to entice the reader to explore the various contributions of this special theme. We also hope to have shed light on how algorithms are imbued with normativities at every step, and how these normativities both shape and are shaped by society. In this manner, the special theme seeks to contribute to a more nuanced discussion about algorithmic normativities in complex sociotechnical practices.

Acknowledgments

We wish to thank all the participants in the network 'Algorithms as devices of power and valuation' for the productive and insightful discussions that have proved indispensable for the production of this special theme. As the name of the network suggests, the network broadly explored the social aspects of algorithms in everyday life, business, government, and science. As the network developed, research about algorithms proliferated, if not exploded, and together we explored the various approaches towards understanding algorithms in society. The network and its collaborative milieu were thus essential for the development and publication of this special theme. Thanks all!

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: *Riksbankens jubileumsfond* for their generous funding of the workshops and PhD summer school that led to this special theme.

ORCID iD

Francis Lee https://orcid.org/0000-0002-7206-2046

Notes

1. See for instance, Amoore (2013); Beer (2009); Dourish (2016); Gillespie (2014); Kitchin (2014); Neyland (2015); Schüll (2012); Seaver (2017); Striphas (2015); Totaro and Ninno (2014); Ziewitz (2017). See also the critical algorithm studies list: https://socialmediacollective.org/reading-lists/critical-algorithm-studies/
2. The articles in this special theme build on conversations spanning three workshops and a PhD summer school on algorithms in society that were held in Stockholm, Sweden, between 2014 and 2017.
3. Also, Haraway's (1992) metaphor of the 'Coyote Trickster' includes non-human actors in a material-semiotic analysis.
4. Classic ANT studies have highlighted for instance how humans 'enroll' microbes, scallops, or speed-bumps in their network to support their various causes.
5. Sometimes in critical algorithm studies, this challenge—to stay above the hood in our playful metaphor—is expressed as going 'beyond opening the black box' to study practice and culture (cf. Geiger, 2017). From an ANT perspective, this interpretation of the black box metaphor is misplaced. It is precisely the manifold and complex relations that the black box contains and relates to that is in focus in this analytical ideal type. Going 'beyond' the black box in this view reifies the idea that studying 'the social' is different than studying 'the technical.'
6. On torque see Bowker and Star (1999).

References

Amoore L (2013) *The Politics of Possibility: Risk and Security beyond Probability*. Durham, NC: Duke University Press.

Angwin J, Larson J, Mattu S, et al. (2016) *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. Available at: www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 3 July 2019).

Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11: 985–1002.

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Burrell J (2016) How the machine thinks: Understanding opacity in machine learning algorithms. *Big Data & Society* 3: 1–12.

Callon M and Law J (1995) Agency and the hybrid collectif. *South Atlantic Quarterly* 94: 481–507.

Chander A (2017) The racist algorithm? *Michigan Law Review* 115: 1022–1045.

Diakopoulos N (2016) Accountability in algorithmic decision making. *Communications of the ACM* 59: 56–62.

Dourish P (2016) Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3: 1–11.

Eubanks V (2017) *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Fourcade M and Healy K (2017) Categories all the way down. *Historical Social Research/Historische Sozialforschung* 42: 286–296.

Galis V and Lee F (2014) A sociology of treason: The construction of weakness. *Science, Technology, & Human Values* 39: 154–179.

Geiger RS (2017) Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture. *Big Data & Society* 4: 1–14.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp. 167–194.

Goodwin C (2000) Action and embodiment within situated human interaction. *Journal of Pragmatics* 32: 1489–1522.

Haraway DJ (1992) The promises of monsters: A regenerative politics for inappropriate/d others. In: Grossberg L, Nelson C and Treichler PA (eds) *Cultural Studies*. New York, NY: Routledge, pp. 295–336.

Kirkpatrick K (2016) Battling algorithmic bias: How do we ensure algorithms treat us fairly? *Communications of the ACM* 59: 16–17.

Kitchin R (2014) Big data, new epistemologies and paradigm shifts. *Big Data & Society* 1: 1–12.

Kraemer F, Overveld K and Peterson M (2010) Is there an ethics of algorithms? *Ethics and Information Technology* 13: 251–260.

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Neyland D (2015) Bearing account-able witness to the ethical algorithmic system. *Science, Technology & Human Values* 41: 50–76.

Neyland D (2018) Something and nothing: On algorithmic deletion, accountability and value. *Science & Technology Studies* 31: 13–29.

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

Nyholm S and Smids J (2016) The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical Theory and Moral Practice* 19: 1275–1289.

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

Pasquale F (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

Roscoe P (2015) A moral economy of transplantation: Competing regimes of value in the allocation of transplant organs. In: Dussauge I, Helgesson C-F and Lee F (eds) *Value Practices in the Life Sciences and Medicine*. Oxford: Oxford University Press, pp. 99–118.

Roscoe P and Chillas S (2015) *Pricing, prizing and praising in the dating services sector – A conceptual approach*. EGOS, European Group for Organizational Studies, Athens, Greece.

Ruppert E, Isin E and Bigo D (2017) Data politics. *Big Data & Society* 4: 1–7.

Sandvig C, Hamilton K, Karahalios K, et al. (2016) When the algorithm itself is a racist. *International Journal of Communication* 10: 4972–4990.

Schüll ND (2012) *Addiction by Design: Machine Gambling in Las Vegas*. Princeton, NJ: Princeton University Press.

Seaver N (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4: 1–12.

Star SL (1991) Power, technologies and the phenomenology of conventions: On being allergic to onions. In: Law J (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge, pp. 26–56.

Striphas T (2015) Algorithmic culture. *European Journal of Cultural Studies* 18: 395–412.

Totaro P and Ninno D (2014) The concept of algorithm as an interpretative key of modern rationality. *Theory Culture & Society* 31: 29–49.

Winner L (1980) Do artifacts have politics? *Daedaleus* 109: 121–136.

Winner L (1993) Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology & Human Values* 18: 362–378.

Woolgar S and Lezaun J (2013) The wrong bin bag: A turn to ontology in science and technology studies. *Social Studies of Science* 43: 321–340.

Zarsky T (2015) The trouble with algorithmic decisions an analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology & Human Values* 118–132.

Ziewitz M (2017) A not quite random walk: Experimenting with the ethnomethods of the algorithm. *Big Data & Society* 4: 1–13.