15 Data aesthetics

Antonia Walford

The age of data is upon us (so we are told). In Euro-America at least, we have been inundated not just with data, but also with images and metaphors of data: from excess, in the form of data floods and deluges, oceans, and cascades (cf. Seaver 2015a), to analogies with natural resources like ‘gold’ or ‘oil,’ to its potential for social change, both positive – as when we are confronted with various data ‘revolutions’ in healthcare or governance – and negative – as when we are told big data analytics herald the end of democracy as we know it (for example Helbing et al. 2017). Disciplines (data science) and professions (data scientist) have grown up around it, and there are industries devoted to its ongoing valorisation through the ramping up of real-time, continuous quantification of any social or natural process, both a continuation of previous historical obsessions with quantification (Porter 1996; Hacking 1990), and a signal of new forms of calculative capital accumulation (Thrift 2005; Zuboff 2018). Data, at least in Euro-America, is emerging as a discrete social phenomenon (Beer 2016), worthy of anthropological attention.

In this chapter, I develop one possible approach to this emergent social phenomenon by asking: what does data *do*, and how does it do it? My contention is that we can begin to answer this through what I am calling an ‘aesthetic’ approach to data, in which aesthetics is understood, not as universal judgements about the beautiful and good, but rather as about what constitutes a ‘persuasiveness of form’ in any cultural or social context (Strathern 1991: 10; cf. Riles 1998). Taking partial inspiration from Alfred Gell (1998)¹ – who, admittedly, eschewed ‘aesthetics’ as a useful term for the anthropology of art – I want to focus on data’s social efficacy, shifting attention from what any given configuration of data *means*, to what it *does*. Although Gell was keen to move away from the culturally universalistic and exclusivist pretensions of traditional and Kantian ‘aesthetic’ approaches to art objects, I maintain the term ‘aesthetics’ in this chapter because to do so is to deliberately invoke an idiom rarely employed in critical discussions of data practices (in a way Gell might have appreciated). A data set is not conventionally considered an art object in the way a painting might be. However, applying this aesthetic approach to data is illuminating because, as many critical data scholars have pointed out, data is rarely discussed as something that is itself made; it is

rather understood as the raw material from which ‘knowledge’ is gleaned. It appears repeatedly in Euro-American business and corporate practice (Coopmans 2014), popular discourse (Anderson 2008) and in scientific discourse (Bowker 2013) as a material or resource out there in the world, which is collected, harvested or mined. It thus is framed as an inert background from which other forms (of knowledge, or action) are elicited. Challenging this presumed *sui generis* status has been the aim of much of the critical work done around data – hence the counter claims that ‘there is no such thing as raw data’ (Gitelman 2013). As these critical scholars have pointed out, data only exists because judgements have been made by humans, who have agendas and presumptions, about what to collect and what not to collect.

But this critique does not go far enough or, rather, it need not stop here. If the argument is that data is not just discovered in the world but actively crafted in different ways, then we need to develop a set of conceptual tools to engage in how this crafting occurs and what different forms of crafting do (see here for example Ratner and Ruppert 2019). My interest in developing an appreciation of data’s aesthetic qualities in this sense is for a more nuanced and rigorous critical engagement with data practices and their effects, that not only focuses on whether data ‘gets it right’ or not. So, although much of social scientific critical work targets the incapacity of big data to accurately represent the fullness of social life (for example, boyd and Crawford 2012), in this chapter I argue that the capacity to accurately represent the world is one of the aesthetic ‘effects’ of data that we need to investigate. I argue that accurate representation (a persuasive form, at least in Euro-America), as one possible aesthetic effect of data, is a result of a process of delicate and committed crafting, of eliciting qualities, forms, patterns, and intensities, and managing their relationship to other forms of social relations (Gell 1998: 9).² This approach will also allow me to countenance the ways in which data’s forms are charismatic, how they capture attention, or command action, in ways other than because of their objectivity or accuracy. The merit in this is that it permits me to sidestep the language of ‘bias’ that is often invoked as a critique of data, problematic because of its corollary that out there, somewhere, is the perfectly accurate data set (cf. Cramer 2019: 37).

A final introductory point to make is that if a study of data aesthetics involves becoming sensitised to the means through which data exerts the effects it does, then aesthetics here also becomes a question of the political work that data does. As Dourish and Mazmanian have argued, a critical study of information needs to pay attention not just to ‘the brute fact of material forms’ of information such as cables, infrastructures or hardware (2013: 4), but to what they describe as the ‘materialities of digital representations’ (Ibid.: 9). That is to say, the organisation of information matters for the effects it can have on people, and the way in which it is used to enact the world (cf. Manovich 1999). As Dourish shows (2014), a relational database can do a kind of political and social work very different to a NoSQL database – in terms of distributing property rights, for example.

In order to begin to flesh out this aesthetic approach to data, I will draw on ethnographic work conducted with environmental scientists and data technicians in Brazil, focusing particularly on the practice of manual data ‘cleaning.’ Data cleaning is understood conventionally as the process by which correct data is sorted from faulty or incorrect data. Focusing on the interplay between revelation and concealment, partiality and holism, and surface and depth that characterise data cleaning allows me to demonstrate how such a practice crafts a data set as a certain kind of internally coherent, persuasive artefact of knowledge with the capacity for particular effects in the world. I subsequently end the chapter by suggesting how an analogous aesthetic analysis could be used to unpack other increasingly pervasive forms of data practice, such as data mining and algorithmic pattern recognition, when machine learning algorithms are trained to ‘see’ patterns in large data sets. Here, the ‘effect’ shifts from that of objectivity, to another form of knowledge that is proving extremely charismatic in contemporary Euro-America.

A feeling for the data

The scientific project I am going to focus on is called the Large Scale Biosphere Atmosphere Experiment in Amazonia (LBA), where I conducted ethnographic fieldwork in 2010. Initiated in 1998, the LBA is responsible for managing, collecting, and making available data from a data collection project that at the time had spanned ten years, with the intention of ascertaining the role of the Amazon forest in the global carbon cycle. It brings together hundreds of different researchers in different scientific disciplines. With help from collaborative partnerships with the US National Aeronautics and Space Administration, and various other Brazilian, European, and North American institutions, the LBA built meteorological towers that stretch up to 20m above the top of the forest canopy, which is almost 70m in places.³ These towers have a profile of equipment on them that measure a number of meteorological variables and generate a large amount of data. This data arrives at the LBA head office in Manaus as an unmarked mass of undifferentiated numbers, which researchers call ‘raw data’ (*dados crus*). It is then put through a quality control process, or ‘cleaned’ (*limpar*), as I was told; another term I heard used was *lapidar*, to cut and polish as one would a gemstone. It is this process of cleaning that I want to concentrate on in this section.⁴

Mindful of Gell’s insight regarding art objects, but also drawing on Annelise Riles’s (1998) analysis of the drafting of UN policy documents by women’s rights lobbyists in Fiji, I want to concentrate on how data cleaning is directed not towards eliciting meaning, so much as towards eliciting the right *form*. Moving away from a critical engagement that seeks to uncover the reality ‘below’ or ‘behind’ bureaucratic instruments like documents, Riles makes a case for understanding these artefacts of modern knowledge through their surfaces and patterns. These documents work, she argues, because they

oscillate between abstraction and concreteness: ‘a simultaneous and mutual apprehension of the document as pattern and the document as an independent object or unit’ (1998: 379). Riles draws this insight out through a comparison with Fijian mat-making: mats lend themselves much more readily to being analysed as aesthetic objects, and she uses them to think in turn about how policy documents are crafted and shaped:

Like mats, we might say, intergovernmental agreements such as the *Pacific Platform for Action* partake in a simple non-representational patterning that is replicated again and again within the document, from one document to the next, and in the mechanics of the conference at which documents are negotiated [...]like the given and self-evident geometric patterns of Fijian mats, moreover, the analytical sequence was all on the surface, dictated, and known at the start.

(Ibid.: 385)

Riles uses this aesthetic approach to illuminate a startlingly new perspective on the efficacy of documents. What seem to be pointless norms and empty flourishes of bureaucratic practice are revealed as integral to the aesthetic of the document, and to the scalar political work the documents do, bringing into view both the local and the global at the same time (Ibid.: 394).

In similar fashion, in this section I unpack the way in which LBA scientific data is crafted by those who work closely with it; and how that crafting then permits the data to have certain effects.⁵ Although Riles’s focus is on the ‘surface’ of documents, what becomes clear is that data is efficacious when its surface form hides a multitude of relations within it.

The data cleaning process at the LBA had a strict protocol which was followed by Raquel and Joanna,⁶ who were responsible for cleaning the LBA tower data. As data cleaning could be quite a nonlinear, complex, and painstaking operation, for the sake of clarity, I have divided it up into three main stages. The first is so mundane as to seem unremarkable initially. It consisted in transferring the raw data – which had arrived on their screens as a block of numbers almost impenetrable to the human eye – into a series of spreadsheets. Spreadsheets, and other tabular forms, are banal and everyday technologies of organisation, but they have important implications for how the data was able to subsequently have effects and travel. The first stage in the data cleaning, then, is when this enormous amorphous body of numbers that is the raw data is copied and pasted into the first spreadsheet table of a nested series of spreadsheets that the data cleaners have on their computers. This spreadsheet has sequential time running down the side, and the different variables the data is meant to be measuring, such as wind speed, rainfall and so on, running along the top. The spreadsheet therefore separates the raw data out into columns of different variables, and times. In order to ensure that the raw data has been copied across into the spreadsheet correctly, it is compared to the first column of the spreadsheet, which is the ‘control’ column, containing

the expected time of measurements by day, month, year, hour, and minute. This is filled out before any data has been copied across. Often, there was a mismatch – for example, the data extended below the edge of the table in the spreadsheet, or the times would not quite match up. One of the data cleaners has what I nicknamed the ‘museum of errors,’ a document with screenshots displaying examples of all the strange permutations that the raw data can arrive in – with missing sections, repeats, and numbers strangely bunched together, or strangely spaced out.

Although it is apparently unremarkable, banal, and executed extremely quickly, this initial transfer of the raw data into the first spreadsheet is very important. The spreadsheet’s structure gives a *shape* to the mass of numbers, differentiating them into columns and rows that have meaning attached to them – specific variables and specific times. It also provides the first clue as to whether something is wrong with the raw data set – which it often is. These deviations are seen as errors in the structure or pattern of the data, *seen as a whole* (Figure 15.1). Casting your eye along the data set, you can see where there might be problems. All deviations from the standard rows and columns have to be corrected, because otherwise, the data will be in the wrong place and, if that occurs, it will be nonsensical. Its meaning depends on its position, on its shape.

Figure 15.1 Raw data before being copied into the spreadsheet.

Source: Photo by author.

The second stage of the data cleaning process is comparing it to what the data cleaners called ‘the range’ – a series of values that have been decided upon for each variable, indicating what is and what is not considered possible for a measurement of that variable. When the raw data spreadsheet is checked against it, the range flags any values that are too high or too low, or when the difference between two consecutive measurements is too great. It is considered to be impossible, for example, for there to be more radiation leaving the surface of the earth than arriving – that is just a ‘physical law,’ I was told, so the ‘reflected radiation’ value could never exceed the ‘incoming radiation’ value.⁷ Other limits could be given by other standard meteorological patterns, the most basic being that there cannot be radiation after a certain time of day (night). The application of these ranges then fixes the limits and boundaries – the outline, as it were – of the data set.

One of the intriguing characteristic of the ranges that the data cleaners at the LBA use is that they are different for different sites of data collection – in fact, as I was told, ‘the range belongs to the place’ (*o range pertence ao lugar*). As a result, the ‘ranges are not fixed. The ranges here [at the LBA] have changed several times.’ The range for one LBA site was worked out using other data sets, standards, and patterns for one site, and then progressively changed to accommodate other sites that were known to be drier, more humid, or have other characteristics. Developing a range is a slow and painstaking process: ‘Experience, literature, meteorological stations – you learn as you go along.’ What comes to shape any incoming raw data set, is another, more trustworthy, data set that has evolved out of years of data collection, and into which the new data set might be incorporated.

This leads me to the third stage of data cleaning, which consists in trying to work out why certain values have been flagged as transgressive. The data cleaners spent an enormous amount of their time working out what was wrong with each data set, and then explaining in a document why they decided to maintain the value as data, or remove it as error. Ascertaining whether the data values are ‘within the possible,’ as they said, proceeds in several ways. There are values that can be quickly dismissed as a gap or an error, because the program that stores the raw data has its own range, which denotes values deemed to be nonsense as NAN (‘Not A Number’). This could be because the instrument lost power, for example. However, some cases are harder to work out – the ‘grey areas’ of possibility, as one researcher put it. For example, there might be a sudden drop in a radiation value, and then a subsequent increase back to what was expected for that time of day in that season. The first way to investigate this occurrence would be to ‘check the relations between the measurements,’ I was told. In this case, could the low radiation measurement be explained by a sudden rain shower, causing a cloud to briefly block the sun’s rays? If so, there should be a spike in rainfall data, an increase in wind speed data, and a change in the pressure data, all at the same time. All these data values would then be compared, and if they demonstrated the right relations to each other then there was good reason to suspect that the value was, in fact, ‘possible.’

Figure 15.2 Revealing the relations using graphs: here, photosynthetically active radiation and short-wave radiation are meant to follow a certain pattern when plotted against each other.

Source: Photo by author.

This comparison was done visually by using graphs (Figure 15.2): ‘there are problems that only emerge when you plot graphs, others that you see immediately’ one of the data cleaners tells me. She goes on, showing me: ‘if you plot this, you can see it’s a problem with the time,’ or ‘here, on the graph these two should show a relation and they don’t.’ If they had a suspicious data value over a period of time, they would graph it against other relevant variables, and ‘look for the relation.’ For example, the latent heat value had ‘passed the minimum’ several times on one occasion, and so ‘when I saw the rainfall spreadsheet, I looked at the same times; when it rains, the temperature drops, and it’s cloudy. For several hours, I saw it was very linked to the rain, through a graph that I plotted.’ She shows me a graph on which she has plotted all the variables, and traces out the lines with her finger, telling me, ‘on this day at this hour, there was a drop in radiation, of short wave, and that’s cloud, and then rain – and that is when the sensible heat values went below the minimum – it’s because of the sensor, do you see?’ Graphs also allowed her to visualise the patterns in the data enough to investigate a suspicious value that has appeared several times. ‘We make a graph of the years and compare them. We know that it rains a lot in São Gabriel [...] when we look at the graph and see that this [the rain data] is really low, if it is sustained then there’s a problem.’ By plotting one variable against another, the data cleaners temporarily reveal the internal patterning within the data. If there is no pattern – if there is too much noise – then there is a problem. This was demonstrated to me one day when one of the data cleaners looked perplexedly at a CO₂ profile graph on the screen and told me the ‘pattern is wrong.’ Even though it was

inside the range, it was not exhibiting the right sort of relation to the time of day, and was therefore suspicious.

The capacity to see these patterns comes with what one researcher called getting a 'feeling' for the data. Going through any raw data they did not have source for, they would remark, 'Oh, that one looks like pressure data,' or 'That one looks like battery data.' On another occasion, one of the data cleaners was going through some old data, and told me 'this one looks like a test, it doesn't look like data.' Like the range, the data from each research site has specific patterns and idiosyncrasies that data cleaners have to learn: 'For example, it's hard to get 100 per cent rainfall there. Here, there's humidity – there, there's dust' and, talking of potential infestations of the instruments: 'It's not the bees there, it's the ants!' However, although data cleaners would sometimes go to the field sites (the towers), they would mostly sit in the office in front of a computer, working on the data. That is to say, what they were getting a 'feeling' for, I suggest, was not exactly how the world works (of which the data would just be an encoded representation), but rather how the world of and in the data works.

The data cleaners are cleaning the data so that it can subsequently travel to other researchers and institutions (although the journey is not always uncontentious, see Walford 2012), as 'trustworthy' data that can be used, manipulated and published. Data from the LBA has appeared in countless articles and book chapters and has been responsible for ascertaining several important features of Amazonian climate and ecological systems, theories that have become associated with a number of very prominent Brazilian scientists. That is, the data has certainly been made to mean something beyond its immediate representational capacity as an index of rainfall or wind speed. But this capacity for the data to travel at all, to enter into relation with other data, depends on the careful crafting work of data cleaners to make the LBA data set into a particular form of knowledge object.

This work entails eliciting, and then concealing again, certain relational patterns from the data sets they are presented with. The data cleaners oscillate between looking at the data set as a whole: Does it have the right shape? And decomposing it into its constituent parts: Do these individual variables demonstrate the right pattern? And then reconstituting it again to check the whole. If the data set is an object, then it is an object that has an internal relationality that can be revealed if necessary. However, once it is cleaned, the necessity to decompose it, to reveal its internal patterning, is removed. The data cleaners are only called upon to demonstrate how the data set relates to itself if there is a question mark over the validity of any of the data values. Otherwise, the data set travels as a knowledge object that is presumed to be fully self-relational – that demonstrates relational integrity to itself. The data cleaners' craft is to produce a data set that has a particular totalising form to it: a form that can be looked at from all angles, and that will appear the same from every perspective. This is a perspectival trick that Bruno Latour calls the 'optical devices' of scientific inscriptions (1990: 43). Does the data set look the

same from all angles? Do the patterns hold however you look at it? The data set becomes trustworthy, not when it can be shown to accurately represent the world, but when it is made into an object that does not allow for any change in perspective – when it in a sense enforces acquiescence. This is therefore not simply a question of standardisation and enforcing a common way of seeing as a skill developed over time (Carusi 2008), but trying to ensure a common way of seeing is in the data set itself for everyone who comes across it. It is an object that can be decomposed into a set of complex self-referential relations, but that in its composed form is singular, unequivocal, and immutable.⁸ What is being crafted so deftly here is the ‘form’ of objectivity itself.

To return to my initial inspiration, Gell insisted that art is to be understood, not as a symbol, but as a form of agency, a translation or transformation of intentions: ‘I view art as a system of action intended to change the world rather than encode symbolic propositions about it’ (1998: 6). Such agency works through what Gell called ‘abduction,’ where the art object becomes an index of social agency – sometimes of the artist or maker – and a form of distributed personhood (Ibid.: 19). This agency is often to ‘enchant’ (Ibid.: 41), to capture attention, to enforce and elicit behaviours, such as when a Kula canoe prow board’s design forces exchange partners to make excessively generous exchanges (Ibid.: 44). As I mentioned earlier, data sets are not conventionally seen as art objects in this sense. But they are crafted with specific intent to persuade others of their veracity, their ability to speak of the world, and this crafting gives them a remarkable power. LBA data cleaning might not seem to evidence the kind of technical virtuosity of a Trobriand prow board carving, nor is the intention of the data cleaner meant to be inscribed into the data set. In fact, data cleaning is assumed to be able to be done by anyone who has the right training, as it is technical work, as opposed to the knowledge production that comes afterwards; data cleaning is not about what the data ‘means,’ and, as a result, data sets do not carry the name of the person who cleaned them in the way that published papers carry the names of those who turned the data into knowledge. But this is aesthetic work, nonetheless, as it is in data cleaning that the data is given the form of its power to persuade. It is in the data-cleaning process, exactly when the data is understood to become ‘true’ and ‘objective,’ that it can come to be what Gell calls a ‘secondary agent’ – that is, capable of carrying people’s (scientists’) agency with it. Scientists can no more be scientists without data as soldiers can be soldiers without weapons (1998: 20), because of what the data allows them to *do*. What is erroneous here is to construe data cleaning as ‘mere’ technical work; it is in fact where the enchantment with and of data begins.

The idea that objectivity is an outcome of what might be glossed as ‘subjective choices’ is not new to Science and Technology Studies (STS) or the Anthropology of Science, and in fact a concern with the representational strategies of science has been at the heart of STS for decades (see Lynch and Woolgar 1990; cf. Coopmans, Lynch, and Vertesi 2014). There has even been explicit engagement with science as an aesthetic endeavour, particularly

around scientific image processing (Lynch and Edgerton 1987; Frow 2014). What I want to draw out of my analysis, however, is that from an aesthetic perspective this is not a question of the subjective versus the objective. The form of objectivity ‘persuades’ others not because of the idiosyncratic subjective proclivities of the data cleaners; in the same way as there is an ‘aesthetic’ to universal knowledge evidenced in its various forms – maps, encyclopaedia, archives and databases (Schaffer et al. 2017) – so too is there an aesthetic to objectivity. Scientific data certainly works in a representational register, and it would be false to suggest that it is not considered valuable because it represents the world more or less accurately. But what I want to emphasise is that it is not *only* representational; or even its capacity to represent in the way it does also depends on its functioning in an aesthetic register, which is non-representational, and in which the data’s form and its relations to itself (rather than to a world ‘out there’) come to determine what it can do, and what (scientific) personhood it becomes instrumental in bringing into being.

Algorithmic vision

To finish, I would like to turn to the question of what this analysis suggests for the future study of ever more-ubiquitous data-related activities that seem very different from manual data cleaning in environmental science, such as machine learning, algorithmic pattern recognition, and data mining. These practices are emerging, not just in commercial spheres as part of new capitalist apparatuses of value extraction (Thatcher et al. 2016), but as forms of governmentality (Amoore 2011), as the bedrock of advances in precision medicine (Prainsack 2017) and neuroscience (Rapp 2015); they are pivotal in the automation of the justice system, of transport infrastructures, and medical diagnosis. That these practices are restructuring society is a claim made quite brazenly by their proponents. As Alex ‘Sandy’ Pentland, an MIT professor recently declared by Forbes as one of the seven most powerful data scientists in the world, states in an interview from 2012:

There are patterns in those individual transactions that are not just averages, they’re the things that are responsible for the flash crash and the Arab spring. You need to get down into these new patterns, these micro-patterns, because they don’t just average out to the classical way of understanding society. We’re entering a new era of social physics, where it’s the details of all the particles – the you and me – that actually determine the outcome.

(Pentland 2012)

The idea that the masses of personal data that is being collected through our communicative and online activities contains within it ‘patterns’ that just need to be brought to light in order to understand the social whole animates most contemporary data science and is at the heart of the current enchantment of

big data. The classic categories of sociological analysis, such as class, race, and gender, are being re-formulated into dynamically shifting ‘micro-patterns’ and ‘vectors’ (Cheney-Lippold 2011).⁹ Important here is an idea that we have already encountered with the LBA data cleaning: the conviction that (big) data is ‘hiding’ these patterns within it. In fact, the aesthetic effect of big data – its persuasive power – derives from the idea that all sorts of things might be ‘hiding’ within it; as Louise Amoore remarked, during a United States House subcommittee hearing in February 2002: ‘The invited panel of experts stated clearly that “our enemies are hiding in open and available information” and that, had surveillance and profiling techniques been in place, the events of 9/11 “could have been predicted and averted”’ (Amoore 2011: 337). Practices of eliciting or revealing these hidden patterns, insights, threats, and so on, function through logics that are not quite recognisable as the ‘objectivity’ of the natural sciences. As Orit Halpern points out, we are confronted with a new objectivity: a ‘new aesthetic and practice of truth; a valorisation of analysis and pattern seeking that I label “communicative objectivity”’ (2014: 15). In another idiom, based on ethnographic work with data analysis in business settings, Catalijne Coopmans has called data mining a form of ‘artful revelation’ (2014); Amoore, for her part, characterises practices employed by private companies employed by the US government to ‘flag’ terrorists in biometric data as working through an ‘ontology of association’ (2011: 27), and the privileging of correlation over causation so infamously claimed for big data analytics is in a sense exactly the celebration of the infinite relationality of big data, its capacity to contain within it countless relational patterns. It is important that these patterns are not on the surface, but concealed. In fact, part of the charisma of these data-driven algorithmic practices is that they are claimed to be so complex that they make decisions based on correlations that humans cannot ‘see’ directly; that what they are doing is ‘rendering perceptible and actionable (almost seeing) that which would otherwise be beyond the threshold of human vision’ (Amoore 2018: 12).

But if at the same time as these machine-learning algorithms seem to make patterns and reveal relations ‘otherwise unavailable to the senses’ (Amoore 2018: 5), they are simultaneously shown to be simply revealing older, equally persuasive, forms. Far from dissolving, classic categories of discrimination are being rendered anew. Racialising and gendering biases, embedded deep into the data sets on which the algorithms are trained, structure these deep patterns, showing them to be nothing more than reflections of human prejudice, and of far older pervasive social patterning that perform and re-perform particular inequalities and violence, particularly racialising ones (Sweeney 2013; McQuillan 2016; Apprigh et al. 2019). PredPol, ‘predictive policing software’ that uses an algorithm to predict where crime will happen based on past crime-rate data, is perhaps one of the most illuminating examples of this. Not only has it come under fire for using biased data sets that rely on data collected disproportionately from racialised populations, and the performative element to its predictive function, which sees these populations

subsequently even more heavily policed; but, more recently, it has emerged that the algorithm used by PredPol was originally designed to predict seismic activity, and several data scientists and mathematicians have come forward to say that it is flawed and too simplistic (Haskins 2019a). Nevertheless, it is heralded by many as the forefront in contemporary policing practice and has been adopted in several US states (Haskins 2019b).

So again we might ask: What do algorithms such as PredPol do when they ‘see’ patterns in data? In her 1993–2008 lectures on Melanesian material culture and decorative practices, published as *Learning to See in Melanesia* (Strathern 1991), Marilyn Strathern makes the pertinent point that we never just ‘see’; people show things to us. This then implies that wherever there is some act of revelation, there are also acts of concealment. When looking at Melanesian ceremonial costumes and masks, her question, ‘What are we being made to see?’ allows her to start to unravel the presumptions woven into what Euro-Americans might think masks ‘do.’ Whereas we in Euro-America might imagine that masks ‘hide’ people, in Melanesia they do the opposite: they display people for the other’s regard, demonstrating their relational potential: ‘When you look at a decorated dancer you see an everted person: what was formerly inside, hidden, is now on display’ (1991: 61). In fact, a person in Melanesia is ‘hidden’ when they are naked, when nothing is being displayed for others to regard, where everything is concealed.

Likewise, we might ask what presumptions of our own shape what we think machine-learning algorithms like that used by PredPol are ‘doing’ when they ‘see’ crime. Whether PredPol accurately predicts crime or not is one aspect of a critical anthropological approach, but it is not the only one we need to be developing. No less important is the means of charting the social and cultural histories and imaginaries that are woven through the dynamics of revelation and concealment that make PredPol so persuasive. We must always keep in mind the labour that goes into these acts of displaying and eclipsing: nothing is just there to be seen.¹⁰ How, despite their inaccuracy, are these new forms of algorithmic objectivity so captivating? How do the patterns they reveal hold together both the recognisable and the not-yet imagined? These are the kind of questions, I argue, that an aesthetic approach to data can begin to answer, and in so doing also provide some means of disruption and resistance.

Notes

- 1 ‘Partial’ because I am not going to follow his schema for understanding the agency of art objects, but would rather make use of his analytical switch from ‘art’ to ‘index,’ whereby he focuses on the ‘social relations in the vicinity of objects mediating social agency’ (1998: 7).
- 2 This is different from, but potentially complementary to, well-established STS approaches to scientific representation and, indeed, to the agency of objects (Lynch and Woolgar 1990; Coopmans et al. 2014).
- 3 The LBA has also recently completed building a 300m meteorological tower.

- 4 There are of course several stages in the production of data that precede this cleaning practice; however, for the purposes of this chapter I am going to concentrate on this specific stage in the data's life cycle.
- 5 This process of data cleaning is of course highly specific to each scientific discipline and sometimes even to each project, so although all sciences will have protocols around data cleaning and curation, this is not intended as a definitive description of scientific data cleaning.
- 6 Pseudonyms are used, at their request.
- 7 Although as these are of different wavelengths ascertaining this is not a simple matter.
- 8 See here Latour's notion of 'immutable mobiles,' although my analysis here deliberately moves away from a dependence on ideas of correspondence between 'sign' and 'referent' that is crucial to Latour's argument (Latour 1999: 24–77).
- 9 See Orit Halpern's book *Beautiful Data* (2014) for a history of these practices.
- 10 My thanks to Rachel Douglas-Jones for her insight here.