Algorithmic paranoia and the convivial alternative

Dr. Dan McQuillan

Department of Computing, Goldsmiths, University of London, UK

Abstract

In a time of big data, thinking about how we are seen and how that affects our lives means changing our idea about who does the seeing. Data produced by machines is most often 'seen' by other machines; the eye is in question is algorithmic. Algorithmic seeing does not produce a computational panopticon but a mechanism of prediction. The authority of its predictions rests on a slippage of the scientific method in to the world of data. Data science inherits some of the problems of science, especially the disembodied 'view from above', and adds new ones of its own. As its core methods like machine learning are based on seeing correlations not understanding causation, it reproduces the prejudices of its input. Rising in to the apparatuses of governance, it reinforces the problematic sides of 'seeing like a state' and links to the recursive production of paranoia. It forces us to ask the question 'what counts as rational seeing?'. Answering this from a position of feminist empiricism reveals different possibilities latent in seeing with machines. Grounded in the idea of conviviality, machine learning may reveal forgotten non-market patterns and enable free and critical learning. It is proposed that a programme to challenge the production of irrational preemption is also a search for the possibility of algorithmic conviviality.

Algorithmic seeing

It is sometimes difficult not to feel that the internet is watching us. Uncanny adverts that appear on websites or next to our social media feeds remind us that someone, or rather something, is attentive to even our most thoughtless browsing. After the Snowden revelations about NSA and GCHQ surveillance, any paranoia we might have had has been re-parameterised; now we know for sure that 'they' are watching everything we do (Electronic Frontier Foundation, 2014). Of all the cultural tropes we can reach for to articulate this experience, one of the most accessible is Orwell's *Nineteen Eighty-Four*. He wrote of screens that watch us back, coupled to a regime that seeks pervasive control over what we think and see. In the year 1984 the internet had yet to achieve speeds of 56 kbit/s but it now seems to be evolving in to something akin to Orwell's dystopian vision, while in most urban conurbations we are constantly over-viewed by CCTV in its various forms. Steve Mann and collaborators use the term 'veillance' to describe this matrix of observation, from the French verb "veiller" which means "to watch" (Mann, 2013). Mann's original motivation was the exploration of alternatives under the banner of 'souveillance', as in 'sous' (from below) rather that

'sur' (from above). With an engineering background, he realised the possibility of creating wearable devices that could watch back, using them as probes to unsettle the asymmetric nature of institutional and commercial video surveillance. In Mann's picture the contestation is between the oversight of the institution and the undersight of the community; a struggle between social formations conducted through mediated vectors of watching (Mann and Ferenbok, 2013). The problem is couched as an asymmetry of transparency, and the social impact as an erosion of privacy. I want to explore how the forms of seeing introduced by big data algorithms differ from these common sense positionings and acculturated ideas of a surveillance society. To start with, I will consider the technical complexity of satellite vision.

Every form of seeing beyond immediate co-presence involves some kind of intervening technology, and this has always modulated the consequences. However, when thinking about the social and political implications of seeing, the technical processes are usually eclipsed by the perceived agency of the operators. The ill-advised invasion of Iraq in 2003 was justified by arguments resting heavily on satellite imagery, where the manipulation and misinterpretation was seen as having come from Colin Powell et al rather than from the technical mechanisms. There is a widespread conviction that satellite images in themselves provide a supra human vision of conditions on the ground, with unambiguous and matchless detail. In fact, satellite vision is a good example of the complexity of machinic seeing, and the way multiple algorithmic translations create the conditions for distortion while claiming the opposite. To reconstitute signals from a satellite sensor in to a meaningful image, there has to be a complicated correction for absorption and scattering by the intervening layers of the atmosphere. Overall, the analogue voltages in the satellite's sensor have to be transformed to digital signals, converted to pixel values, algorithmically corrected for atmospheric effects using optical characteristics from other measurements or models, assembled in to an image, stored in a standard format and transmitted by error-correcting and signal processing protocols (Fallah-Adl, 1995). As Susan Schuppli points out, satellite imagery is "mobilised as indexical truth claims" but "politics enters into the visual field not simply at the level of representation - the content displayed in the image - but at the structural level of its information acquisition, processing, and transmission" (Schuppli, 2013). Satellite vision shows how high-tech prostheses that are assumed to usher in unparalleled transparency are actually introducing additional strata of manipulation. It is a similar form of machinic politics that we must come to question when considering big data's algorithmic vision.

The emergence of big data prompts us to consider the how algorithms see. The signals that give algorithms their sight are drawn from our myriad entanglements with technical systems; not just the

internet, but every other data-emitting infrastructure that scatters data as a response to our passing. Big data seems to deliver a datafied X-ray vision that does not distinguish between public and private realms, and is thus perceived as an amplified threat to privacy (Lyon, 2014). However, rather than consider this as an extension of previously existing veillance, I will concentrate on the uniquely distorting effects of the processing that occurs at the structural level. Whereas the power of surveillance in Orwell's vision depended on human watching facilitated by the transparent portal of the vision screen, big data is processed in to meaning by machines; specifically by datamining and machine learning algorithms (Hastie et al., 2003). To understand the seeing produced by big data we should grasp the nature of this algorithmic eye, through an appreciation of the way the operations of datamining and machine learning modulate the process of meaning making. As with human vision, machine learning algorithms attempt to structure a vast and changing input in to recognisable patterns. The assumption is that there is some underlying function which can relate input data to the targeted output (that which you want to be able to 'see'). This function draws on a set of features present in the input, which are processed by machine learning algorithms such as decision trees (Yee and Chu, 2015), k-means clustering (Piech and Ng, 2012) and artificial neural networks . A basic model for this method of converting input data to a pattern is fitting a straight line to a set of scattered points using the method of least squares. Finding the line that minimises the distance from each point to the line allows you to see it; to assert that, whatever the scattering of the original points, you have revealed the deeper relationship between them. The parameters of your line allow you to predict a y (an 'outcome') for a given x (a 'feature'). In big data, the line is not fitted to x-y coordinates but in a multidimensional space reflecting the large number of features that the algorithm is trying to combine. After repeated fitting to training data, the algorithm 'learns' to generalise its predictions for new cases. Thus, big data algorithms try to reveal regularities in the data which can be used to make predictions. But this algorithmic analysis rests on assumptions that potentially cloud the resulting vision.

The process of fitting a line to a set of points is a fundamental operation of science; a tool to uncover the underlying regularities of nature that we believe to be scientific laws. Science aims to reveal the casual mechanisms that lead to observations. Big data algorithms, on the other hand, are not attempting to reveal causal mechanisms but simply to relate the pattern of past observations to the prediction of future observations. Essentially, they substitute correlation for causation. The power of big data algorithms comes from being able to do this with vast patterns and even vaster data sets. Predicting the likelihood of someone making a particular online purchase may involve hundreds of correlated features (a browsing history of hundreds of urls, the browser used, location of the user, time of day, weather conditions, their friendship networks on social media, and so on)

and millions of training data points from other people's previous purchases (Duhigg, 2012). It seems that big data algorithms allow us to see previously invisible connections. But the very bigness of big data introduces an inherent opacity. In all but the most simple cases, it is not possible to directly apprehend how a machine learning algorithm has traversed the data because of the number of variables involved and the complexity of the function that the algorithm has derived to map inputs on to output. We simply can't see how it works, we just have to acknowledge that it has produced some statement of likelihood about a future state. As pointed out in a previous issue of this journal, there is no human interpretation of why a machine learning algorithm (in this case, a Support Vector Machine) should consider words like 'visit' or 'will' as indicating spam, alongside more obvious words like 'money' or 'contact' (Burrell, 2016). The operations of machine learning preclude openness and, by implication, accountability.

The algorithmic eye is not ocular but oracular. It is the eye of the seer, peering in to the future to produce predictions that demand both interpretation and action if we are to avoid misfortune. The modulation of the present in the name of an algorithmic vision of the future forces us to ask what elements from the past are being projected in to that future, and hence in to the now. We can recall that one of the oldest pieces of computing jargon, dating back to the 1960s, is 'Garbage In, Garbage Out'. The strength of big data algorithms is often seen as their ability to derive meaning from vastly diverse forms of data, but that doesn't make them immune from distorted patterns in the input (NYU School of Law, 2016). If a predictive policing algorithm relies on police crime records, as it almost certainly must whatever else it ingests, then it risks reproducing any patterns of prejudice that may be latent in that data. For example, research has shown that the data on sentences of black people in the USA doesn't only reflect the severity of the crimes but the prejudice of police and prosecutors: "Black arrestees in the federal system - particularly black men - experience moderately but significantly worse case outcomes than do white defendants arrested for the same crimes and with the same criminal history. Most of that disparity appears to be introduced at the initial charging stage, which has previously been overlooked by the literature on racial disparity in criminal justice" (Rehavi and Starr, 2012). Evidence that this kind of bias is already being mobilised through data science comes from a recent analysis of the COMPAS recidivism algorithm, showing that it discriminates against black defendents (Larson, J. et al., 2016). Algorithms may also introduce distortions of their own; weightings derived from their combinatory powers that make mathematical sense in context but would be interpreted as socially unacceptable if set against principles of fairness and equality. The challenge comes from the concealment of potential prejudice by the inherent opacity of the process.

We can grasp the tendency towards opacity and potential prejudice if we consider in more detail how machine learning algorithms actually function. Fitting a line to a set of scattered points is one example of a regression algorithm, where the assumed relationship between a set of features x (e.g. property location, number of rooms, distance to local school) and a target output y (property price) is modelled by a function f, so y = f(x). The algorithms make the least bad guess at the function by minimising the cost of the difference between the outputs predicted by the function and a set of known outputs called a training set. The cost J is defined as a suitable way of penalising the predictions of function f for being different from the observed values and this cost varies as a function of the parameters θ of the fit, so the cost is a function J(θ) not of the features but of the parameters (i.e. of the relative weights of the different features). There are different ways to minimise the cost function J; in gradient descent, for example, the calculation iterates towards the minimum point like a blind mountaineer descending step-by-step in to a crater (the difference being that this crater is many-dimensional). This is how an algorithm 'learns': something of interest to people ('what's the pattern here?') is translated in to something computers are good at (doing thousands of calculations quickly to minimise a function). A similar process can be made to work for classification (e.g. terrorist / not terrorist) by constructing function f as a probability function and J as a cost function that strongly pushes the outputs to one or other classification (this is known as logistic regression). The derived algorithm can be used to calculate a risk score for any new person of interest.

So far, so mathematical. There is process at work which echoes that of mathematical physics; the abstraction of the world into equations in search of hidden order. That the scientific order should be mathematical can still be a source of wonder: see for example 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences' by Nobel prize winner Eugene Wigner (Wigner, 1960). But the point here is the relatively arbitrary character of machine learning, which is driven by both mathematical and cultural factors. Mathematically, the algorithm will force a least bad fit whether there's any causal connection or not. This fit depends on the choice of cost function and minimisation process, and may be 'biased' (which in machine learning terms means forcing a fit to the wrong kind of curve). The result may also be distorted due to 'overfitting'; imagine a very wiggly line connecting a set of more-or-less linear dots – while the line fits the observed dots exactly it will clearly be a poor predictor of future observations. And while the mathematics is precise, if potentially misleading, the process itself is malleable. Unlike physical quantities like mass or energy, machine learning's features don't rest on a wider body of experimental dependencies; their construction is fairly unconstrained. Feature engineering is the selecting and

construction of features and is one of the dark arts of machine learning. It draws on the intuition and motivation of the engineer or computer scientist which, while based on their understanding of algorithms, will also be informed by their world view and the corporate or governmental culture in which they operate. While machine learning algorithms are debated in scientific papers, real world instances of their application are generally not subject to the grounding effect of peer review; rather, they are closely guarded commercial or political secrets. The people acting on their final recommendations usually have no idea about all this mathematical or interpretational flexibility, and simply take the results as given.

The algorithmic gaze is part of a cybernetic system that acts on its own predictions. Seeing the future, it preempts it. As we have seen, there exists the potential, indeed likelihood, of algorithmic bias which violates the concept of fairness and equality. But the primary victim is due process; while the action is immediate the reasoning is inaccessible and immune to scrutiny. "Big data enables a universalizable strategy of preemptive social decisionmaking. Such a strategy renders individuals unable to observe, understand, participate in, or respond to information gathered or assumptions made about them. When one considers that big data can be used to make important decisions that implicate us without our even knowing it, preemptive social decision making is antithetical to privacy and due process values" (Earle and Kerr, 2013). Algorithmic vision is erasing the due process enshrined in law since Clause 39 of the Magna Carta (The Magna Carta Project, 2015). How has the computational manipulation of data risen to a point where it can undermine basic juridical principles?

Algorithmic vision derives authority from its association with science. The idea that computers can make mistakes is commonplace, but the new methods characterise themselves as akin to science by operating under the banner of 'data science'. The apparent rigour of the computational mathematics casts the oracular pronouncements with an aura of neutrality and objectivity, which can be used to defend against the critique that they carry any social prejudice (Gorner, 2013). If algorithmic vision is a way of seeing, is it scientific? Certainly, many advances in science have corresponded to new ways of seeing, from Robert Hooke's 17th century microscopy to the latest enthusiasm for gravitational waves. And science, like data science, is seeking repeatable regularities. The root power of science comes from the concept of a generalisable law that predicts physical phenomena, and the grail of any machine learning is prediction that generalises to new cases. Moreover, the idea that computer modelling can substitute for physical experiment goes back as far as the origins of the Monte Carlo method in the nuclear calculations of the Manhattan Project (Metropolis, 1987). But even where data science adopts whole mathematical models from science, as in predictive policing

company Pred Pol's use of earthquake models to predict petty crime (Twachtman, 2013), it is glossing over its own lack of groundedness. This hubris leaves data science open to be a vector for social consequences that in other contexts would be challenged as ideology. The risk is that the statistical regression at the heart of machine learning will become an engine of forms that are socially regressive, in the same way that a pseudo-science like Craniometry flourished in the era of colonialism (Morton and Combe, 1839). We must generalise our own questioning of big data algorithms to ask what kind of society will be produced by this new apparatus. Having looked at the ways machine learning is both like and unlike science, we should look in the other direction and ask about its social resonances. What forms of governance and government resonate with algorithmic seeing, and what social distortions may result? We can parameterise the possibilities by drawing James C. Scott's text 'Seeing Like a State' (Scott, 1999).

Algorithmic paranoia

While endeavouring to understand the apparent hostility of many nation states to the nomadic elements of their own populations, Scott came to see state actions in the wider frame of attempts to make society legible. Essentially, the state sought forms of subjectivation that 'simplified the classic state functions such as taxation, conscription, and the prevention of rebellion'. Scott describes the pre-modern state as 'partially blind', whereas the modern state transformed illegible and messy social practices into constrained forms that can be seen and therefore read. In his book, Scott's particular focus is the production of disasters in the name of development; historical and contemporary mega-projects of the state in the name of the common good that instead produce human suffering at scale. He had in mind the Great Leap Forward in China, collectivisation in Russia, compulsory villagisation in Tanzania, Mozambique and Ethiopia, and more contemporary development which is "littered with the debris of huge agricultural schemes and new cities (think of Brasilia or Chandigarh) that have failed their residents". In attempting to understand why, as he puts it, so many well-intended schemes to improve the human condition have gone so tragically awry, Scott proposes a four-fold structure for these social engineering disasters. These elements are: the bureaucratic ordering of nature and society, a high-modernist ideology, an authoritarian state, and a civil society that lacks the capacity for resistance. I will argue that systems of algorithmic governance instantiate all four elements of Scott's structure; in other words, all the necessary elements for things to go tragically awry.

Clearly, big data seeks to bring about a new form of ordered legibility of society according to its

own logic of patterns. Big data also corresponds to Scott's high modernism, by which he means a confidence in a mastery of nature and a rational design of social order that is ideological but borrows its legitimacy from science and technology. As I outlined earlier, the term 'data science' embodies this muscular self-confidence in the propellant nature of new numeric insights to bring unprecedented progress in industry, commerce and social order. Algorithmic action is authoritarian in the sense I have also outlined, in that it eludes democratic oversight and, so far, evades a social discourse capable of challenging its teleology. The fourth factor, that Scott calls 'a prostrate civil society', comes exactly because the consequences of algorithmic action are not yet conceived in forms that make them legible to traditional social actors, nor in ways that allow us to think of a social counter power. The potential for such a counter-power, in the form of convivial alternatives, is a point I will return to in the last part of the paper. Suffice it to say that algorithmic governance in some sense unites Scott's four preconditions for state-led developmental disaster. This worrying tendency is accentuated by a further association with the secret state.

Judging by the Snowden revelations, the secret state has already found a friend in big data. While the slides he leaked are unambiguous about the way spy agencies hoover up all the data, it is less clear what happens to it afterwards and how exactly it is analysed. Nevertheless, the role of machine learning in secret state machinations can be glimpsed from the description of SKYNET. SKYNET is a programme for courier detection via machine learning (The Intercept, 2015b). The question posed by the leaked SKYNET slides is "Given a handful of courier selectors, can we find others that behave similarly by analysing GSM metadata?". This algorithmic hunt for al Qaeda uses Dialled Number Recognition data, such as time, duration, who called whom, and user location, to analyse cellular network metadata of 55 million people. The slides brag that certain actions, such as frequently turning off the handset or swapping SIM cards, are interpreted as attempts to evade mass surveillance. The assumption is that the behaviour of terrorists is sufficiently different to ordinary citizens that pattern-finding algorithms (in this case, Random Decision Forests) can filter them out (The Intercept, 2015a). This method of pattern finding is based on Decision Trees, where the training data is split into two based on the value of a feature (e.g. 'is age < 24?'), then each child node in turn is split on another feature, and so on and so on, until at the bottom of the tree each descendent represents a value of the target variable (in this case, 'terrorist courier / not terrorist courier'). The cumulative decision for each final answer can be read off as the series of yes/no decisions made on the path from the top. Any new input is classified by passing it down the tree so that it ends up being assigned a target value. The Random Decision Forest improves this by creating lots of complementary trees from slightly randomised sets of data & decisions. However, this method is still known for overfitting where the data set is particularly noisy.

In this case, the available training set of known terrorists consisted of only seven examples. Six were used to train the algorithm and the seventh was used to test how well it worked. The outcome was used to derive statistics about false negatives (terrorists who would be missed) and false positives (ordinary citizens who would be targeted). Under any circumstances, even the smallest rate of false positives quoted on the slides would lead to thousands of innocent people being categorised as terrorist couriers. A stronger critique of the slides by statistical and machine learning specialists suggests that the reasoning used is flawed even in data science terms because of the asymmetric scale of the known terrorist subset compared to the population, the character of the specific algorithms (their tendency to overfit), and the lack of resulting quality indicators for the method (Grothoff and Porup, 2016). Their conclusion was that actions based on the reasoning in the slides would be far worse than estimated. Although the leaked slides don't reveal how directly the machine learning results are used to identify drone targets, former head of the National Security Agency Gen. Michael Hayden admitted during a discussion at a university symposium that "We kill people based on metadata" (Ferran, 2014).

It is perhaps easy to understand how in the disembodied world of drones, whose operators sit in cushioned and air-conditioned comfort half a globe away, the sideways slippage of the concept 'false positives' from the innocuousness of spam email filters to the finality of extrajudicial killing might occur with only some delayed Post Traumatic Stress Disorder as its trace (Costello, 2015). But the in the increasingly closed loops of algorithmic governance and the emergence of 'smart' social policy, a similar distancing will come in to play in all areas of social reasoning. That the NSA & GCHQ have an insatiable appetite for data fits the simple veillance model. But intelligence agencies, by their nature, see threats everywhere. Through machinic lenses, the general mode of governance may move from 'seeing like a state' to 'seeing like a secret state'. The adoption of algorithmic seeing condenses Scott's four-fold criteria of state vision with a fifth: 'paranoia'.

As we have seen, the form of legibility added by big data algorithms is prediction; reading the population as a form of reading the runes. Moreover this prediction is linked to preemption, a doing that follows directly from the seeing. We can understand how this can lead to a mode of operation that emulates paranoia by considering the idea of perception, or rather apperception. As originally used by Leibniz, the term apperception simply meant the coming in to consciousness of small, unconscious perceptions (Kulstad and Carlin, 2013). There is a particular irony here in the link to Leibniz, who developed his symbolic logic from a belief that human reasoning could be translated to calculations, the solution of which could resolve human conflicts. While his ideas of reasoning

through symbols and calculations can be seen as an ancestor of the Turing machine and hence of general computation, the specific forms of computation we are considering here have very difference consequences to the ones that Leibniz hoped for. Since his time, psychology has developed the term apperception to mean the assimilation of new observations or experiences in to the totality of past experiences. We can ask what counts as past experience for a machine learning algorithm, and how is this assimilated. The predictive algorithms that are now at play are mobilised by situations of risk; financial risk, social risk or security risk. Likewise, the psychological state of paranoia is a thought process that is heavily influenced by threat and anxiety. In most cases, the decision-making process of machine learning cannot be translated in to human reasoning; they are, in that sense, irrational, as is the paranoiac mode of thought. What most people would see as coincidence a paranoid person may believe was intentional, while the whole of machine learning is based on finding meaning in patterns of coincidence. Paranoia and machine learning weave around each other like the serpents around Hermes' staff. But what really drives the resonance of big data analytics and paranoia is the tendency for recursion; for the machinic processes to feed on themselves.

Mackenzie highlighted the problem of the perfomativity of prediction in machine learning. That is, while the assumptions of machine learning assume stable classifications to be discovered, the actions of predictive algorithms may themselves change people's behaviour in ways that the model did not learn about when it was trained (Mackenzie, 2015). While he would see this as having the potential to thwart the process of datamining, I want to emphasise the potential for a feedback and recursion. Through preemption, the next instantiation of algorithms is going to learn from conditions influenced by the prior ones. As Massumi pointed out, preemption 'brings the future in to the present' (Massumi, 2005). Thus, predictions change the conditions for the learning of machine learning. The situation has the potential for a machinic form of paranoiac self-justification, an algorithmic attribution bias that generates systematic errors in evaluating the reasons for observed behaviours. Under these conditions, algorithmic seeing tends towards paranoia.

In fact, recursion forms the basis of a whole class of algorithms that learn for themselves, in the form of deep learning neural networks. Neural networks in general are an interestingly different form of machine learning. They have a long and mixed pedigree, but are popular because of their particular suitability for big data. Neural networks consist of layers of 'neurons' that are simple computational nodes. Instead of the usual programming paradigm, where data is fetched from memory and serially processed according to a set of instructions, neural networks are parallel structures where the connection between input and output consists of the weighted interconnections

of intervening nodes. Ordinary neural networks are the same as other modes of machine learning in that they have to be trained. In their case, the learning doesn't come through a formal algorithm like K-means but by 'back propagation', where a signal passed back to the intervening layer is derived from the difference between the current output and the desired output. This alters the weights of the connections between those intervening layers, and iteratively improves the neural network's ability to reproduce the training data. They're called neural networks because it was imagined that this way of working models the operation of neurons in mammalian brains, where the production of meaning comes from the million- or billion-scale firings of simple entities that respond to incoming electrical signals by firing off more signals to all the other neurons they are connected to. The origin of neural networks stretches back to the 1940s when Warren McCulloch and Walter Pitts, influenced by Leibniz's logic and contemporary efforts to model biology as a mathematical science, wrote their seminal paper 'A Logical Calculus of Ideas Immanent in Nervous Activity' (McCulloch and Pitts, 1943). The link to 'seeing' was made manifest by Frank Rosenblatt's Mark 1 Perceptron machine in the late 1950s, which physically implemented his perceptron algorithm for image recognition by connecting to a 20×20 array of cadmium sulfide photocells to produce a 400-pixel image (Rosenblatt, 1958). The hope was that neural networks would bypass the limitations of ordinary programming and imitate the incredible learning powers of animal brains.

It turns out that the workings of organic brains may not be represented very closely by artificial neural network works. But nevertheless, neural networks have a particular facility for tasks where the desired output is hard to parameterise, but where there are plenty of examples. A typical application is handwriting; it's hard to write a formal description of all the different ways people write a particular letter, but given enough examples, neural networks get closer than other machine learning algorithms to the human ability to identify the letters despite all the natural variations (Schmidhuber, 2013). In neural networks there is no explicit algorithm to examine; the learning is represented by the state of the intervening layer. For a number of reasons, but especially because of available computing resources, neural networks were originally developed to have one or two intervening (hidden) layers between input and output. But the scaling of computational power, particularly by entities like Google, has fuelled an interest in neural networks that do deep learning. That is, where there are many hidden layers between input and output. Each learns from the one below, and develops a successively more abstract representation of what is being learned (Goodfellow et al., 2016). This deep learning reproduces an aspect of what is understood to happen in real brains, in that there are successive layers of abstraction and thus meaning formation. When applied to pattern recognition, for example, initial layers might be learning to recognise edges or corners, while deeper layers might be learning how to assemble these in to representations of 'face'

or 'car'. The real excitement about deep learning neural networks comes from their potential to learn for themselves. Rather than requiring the painfully careful preparation of pre-categorised training data, they can simply be force fed a large number of data sets to learn from.

Google engineers attempting to understand how their deep learning networks are learning to recognise images, for example, have taken snapshots of the representations from each hidden layer. These are a fascinating insight in to how a neural network 'thinks', as it is possible to see the gradual emergence of a blurred order in the contents of successive layers. In an experiment to bring out the latent content of these layers, the Google engineers used the snapshots as fresh input, meaning the algorithm amplified its own partial perceptions. They labelled this process 'inceptionism' (Mordvintsev et al., 2015). Applying this to images produces surprising results; depending on the input, the artificial neural network finds faces within faces within faces, or combines its partial idea of various animals in to dog-pigs or fish-camels. The algorithm's 'wishful thinking' produces a surreal aesthetic of apophenia, where apophenia is 'the perception of connectedness in unrelated phenomena' along with 'a sense of abnormal meaningfulness' (Brugger, 2001). The type of apophenia involving images or sounds is known as pareidolia, and the products of Google's experiments are a kind of fractal pareidolia.

While the engineers see inceptionism as a way to visualise how neural networks work and thereby to improve network architecture, I propose that these images are a figure for a future experience of life under algorithmic governance. These neural networks accentuate the algorithmic characteristics that I have described as contributing to the potential production of injustice. Their reasoning is obscure and irreversible, they will find patterns even where those patterns have no meaning, and they will feed on themselves. When applied to imagery this can be either constrained or, in a research context, enjoyed for its own aesthetics. Let loose on the messy realm of the social, this amounts to the production of computational conspiracy theories. The future realm of algorithmic social order may be a habitus prehended by Kafka's The Trial (Kafka, 2010). Not only because of being a form of life that is interrupted by mysterious juridical interventions but also because of the psychological rendings that we subjects, like The Trial's suspect Joseph K., will inflict on ourselves as a response. While the frame of Kafka's tale is the matter-of-fact surrealness of his arrest and arraignment, a large part of the text is the protagonist's almost sado-masochistic response to the trial's reshaping of his social relationships and social status. How will our mental models of normal relations between ourselves, others and institutions be distorted if subjected to apparently baseless interventions stamped with the authority of calculative objectivity? Like the way the inceptionism imagery blossoms with semi-familiar figures filling fractal gaps, Joseph K.'s world erupts with

events & structures dislocated from their familiar settings, like the court room nested within the dank corridors of a deprived housing estate. We should not be surprised if a world where deep machine learning shapes significant interactions turns out to be steeped in parallelised paranoia.

Algorithmic conviviality

How can we find our way out of the looking glass world that predictive algorithms could bring about? I don't believe it will suffice to point out the social cost of false positives, as the commitment to this mode of operation runs too deep. Unlike Athena, algorithmic rationality did not spring fully formed from the head of Zeus. It is built on the foundations of science and of computation and called forth by a society whose value system is ultimately calculative. Big data algorithms are a way of looking at the world, and as Tibetan Buddhist Sogyal Rinpoche points out, 'how you look is how you see' (Rinpoche, 1996). To expect that the consequences will be abated by regulating certain actions based on certain forms of pattern finding is like trying to hold back the tide. The challenge is to find a different vision, a different centre of gravity, that will produce an alternative unfolding of these new modes of perception.

We are seeking to recover difference from the descendants of the Difference Engine. While we should be unsettled by the emerging problems of algorithmic apophenia, we could pivot our perspective to say, like Pasquinelli, that creativity and paranoia share a perception of a surplus of meaning, so apophenia could equally be about the invention of a future out of a meaningless present (transmediale, 2015). Mackenzie acknowledges the dark side, saying "Almost everything we know about the historical experience of action, freedom, collective becomings or transformations points in a different direction to the technologies themselves" (Mackenzie, 2015), but he also sees the possibility of slippage in the unstable performativity of machine learning and asks "Could the production of prediction also increase the diversity of social production or inform new collectives?. Since it is pre-individual in focus, the 'unknown function' that generates the data might also diagram different forms of association".

How can we recover something truly different from the paranoid predictions of algorithmic governance? I will outline two steps towards a position that has some traction for change. The first is to deflate the scientific and empirical hubris that, through the mechanism of prediction, is channelling the myth of progress in to preemptive interventions in the present. The second is an inversion of tools like machine learning to serve a purpose that Ivan Illich described as 'convivial'.

We need to refocus our vision. As Donna Haraway said, "Struggles over what will count as rational accounts of the world are struggles over how to see" (Haraway, 1988). It is to her ideas about feminist empiricism that I think we can turn as a brake to the acceleration of algorithmic authority, whose momentum comes from claims to objectivity. It is no accident that concerns about algorithmic authority resonate closely with Haraway's wider critique of science. Her paper on 'Situated Knowledges' identifies the primary toxicity of orthodox scientific interpretation as being about 'the god trick of seeing everything from nowhere'; in other words, a universal and objective vision from above. Indeed this is what the latest wave of quantified pattern-finding also lays claim to; an insight from above, a point of privilege that sees distributed connections invisible to us mere Flatlanders (Abbott, 2015). It derives authorization for subsequent interventions based on an objectivity that out-scales alternatives. But we have seen the obscurity of its decision-making and the difficulty of challenging what we cannot convert to human reason. Thus, big data recapitulates the trick that Haraway criticises; claiming objectivity while avoiding accountability.

Instead of responding to patriarchal science by dropping the idea of a faithful account of a real world, Haraway promotes the alternative of an embodied objectivity, a situated knowledge that accepts the limits and responsibility of seeing things from a particular point of view. In her view, and in the wider field of feminist and critical empiricism, this recognition that all viewpoints come with social, cultural and metaphorical baggage can claim a stronger objectivity against a transcendental view that denies its own history and contingency. For Haraway the idea of vision and seeing is a way to unpick the current state of affairs, and fortunately for us this maps straight on to our need to question the machinery of seeing. "Embodiment is significant prosthesis; objectivity cannot be about fixed vision when what counts as an object is precisely what world history turns out to be about...The visual metaphor allows one to go beyond fixed appearances, which are only the end products. The metaphor invites us to investigate the varied apparatuses of visual production, including the prosthetic technologies interfaced with our biological eyes and brains... It is in the intricacies of these visualization technologies in which we are embedded that we will find metaphors and means for understanding and intervening in the patterns of objectification in the world-that is, the patterns of reality for which we must be accountable” (Haraway, 1988).

I conclude that feminist empiricism is a necessary restraint against the tendency for the apparatus of algorithmic vision to exploit ideas of objectivity to seek power without accountability. One of the core strengths of Haraway's work is that she critiques the colonial characteristics of science without abandoning a positive empirical project. Similarly, I will consider the possibilities for machine

learning as a mode of care and nurturing despite the accruing evidence of its corrosive power. I will do this through the work of Ivan Illich, in particular his 1973 text 'Tools for Conviviality', by using it to make the case for convivial algorithms (Illich, 1975).

Illich chose the term convivial to designate a modern society of responsibly limited tools. The attribute 'convivial' is deliberately intended for the tools themselves rather than the social forms, but the definition of tools is broad. He was trying to articulate the effect of tools in the triadic relationship of persons, tools and collectivity. Tools, for Illich, means something that readers of Foucault might label apparatus or assemblage; most definitely hardware and machines, but also productive institutions that produce tangible commodities, and productive institutions that produce intangible commodities like health and education. Through the notion of tools he intended to sweep up all forms of engineered instrumentality to examine their role in denying conviviality. The 'conviviality' that Illich promoted is a lively term that is, naturally enough, hard to pin down to a dictionary definition. But we can immediately assert its relevance when considering algorithmic governance, because it describes "autonomous and creative" activity "in contrast with the conditioned response of persons" (Illich, 1975: 24). As we have seen, the emerging algorithmic apparatus acts as a megastructure of conditioning, controlling our responses by its own response of excising certain possibilities before they have a chance to manifest. Machine learning in a control society recapitulates the things that Illich was most criticising and the conditions he sought to escape from. His means of escape was one we are quite unfamiliar with at the moment; the idea of limiting our tools. "A convivial society should be designed to allow all its members the most autonomous action by means of tools least controlled by others. People feel joy, as opposed to mere pleasure, to the extent that their activities are creative; while the growth of tools beyond a certain point increases regimentation, dependence, exploitation, and impotence" (Illich, 1975: 34). In his eyes, the unlimited growth of tools was a driver of negative consequences and the new goal should be appropriate limits, to maintain overall balance.

Illich talks about the idea of negative design criteria to define the limits within which tools are kept. The idea of limits is a lever for inversion; for inverting the immiserating impact of tools and institutional structures towards forms of creative and autonomous interrelationship. He saw a convivial society as providing freedom through interdependence, where convivial retooling can "guarantee for each member the most ample and free access to the tools of the community and limit this freedom only in favour of another member's equal freedom" (Illich, 1975: 25). Illich had a particular focus on the idea of justice as a pillar of the convivial society, and felt it was vital to limit the scope of tools in order to enable justice to come forth. His emphasis was not on centralised

juridical mechanisms but on a broader empowerment. "A methodology by which to recognize when corporate tools become destructive of society itself requires the recognition of the value of distributory and participatory justice" (Illich, 1975: 31). But he didn't define justice as a pristine value outside a complex world of power relationships mediated by operant tools and structures. Rather, he said that "rationally designed convivial tools have become the basis for participatory justice". We can say, therefore, that the way to preserve justice under current conditions is not simply to defend existing structures but to proactively seek the positive diffusion of justice through our new tools. Whether this would involve specific forms of machine learning or perhaps, as was recently proposed, the blockchain (Buterin, 2016), is yet to be determined.

The method by which limits are to be explored Illich called 'counterfoil research' whose goal is to detect "the incipient stages of murderous logic in a tool" and to "devise tools and tool-systems that optimize the balance of life, thereby maximizing liberty for all" (Illich, 1975). His idea of counterfoil research was articulated at a similar time and in a similar spirit to Paulo Freire's idea of critical pedagogy (Freire, 2000). Both sought to valorise the 'view from below' and wrest decision making from the distant realm of experts who are assimilated in to the logic of the current system. This both preempts the notion of souveillance and is different to it, through the emphasis on the active production of human flourishing. Because of Illich's specific focus on tools, his thinking speaks directly to the problems of algorithmic prediction that have been described in this paper. "Counterfoil research must clarify and dramatize the relationship of people to their tools. It ought to hold constantly before the public the resources that are available and the consequences of their use in various ways. It should impress on people the existence of any trend that threatens one of the major balances on which life depends. Counterfoil research leads to the identification of those classes of people most immediately hurt by such trends and helps people to identify themselves as members of such classes" (Illich, 1975: 98). We can see the stirrings of counterfoil research in emerging critical work on algorithms, which attempt to both identify and ameliorate the classes of people who might be hurt by algorithmic prejudice (UNews, 2015). However, what this work lacks by comparison to Illich is the connection to wider vision of a better world through recalibrating our relationship to our tools.

Conclusions

In this paper I have explored the question of algorithmic vision as a way to understand the

consequences of ceding social power to predictive analytics. In considering vision, I argued that familiar ideas of centralised surveillance and privacy violation are not sufficient to explore these consequences. Rather, I tried to outline the machinic characteristics of big data seeing and its potential to produce distortion and prejudice, driven by a fusion of mathematics and culture. From machine learning arises a mathematico-cultural force with relations to wider systems of power. By considering the association with science that lends authority to these mechanisms, I aimed to go beyond the specifics of algorithmic bias to consider the consequences at a social level, through the idea of 'seeing like a state'. This suggests there is a strong possibility for algorithmic governance to produce bad outcomes at scale, whether or not they are intended. Moreover, the operation of algorithmic governance tends towards 'seeing like a secret state', with it's own apperceptive paranoia. Through Kafka's figure of Joseph K, I asked what it might be like to be a subject of such a state of affairs.

Seeking a way out of these enclosures, I suggested that we can reclaim a more accountable way of seeing through feminist empiricisim, as articulated in particular by Donna Haraway. I propose that this be made concrete through Ivan Illich's idea of convivial technology, through the ideas of limits, negative design and counterfoil research. The direction of travel of algorithmic prediction and preemption under current conditions is to close down possible futures, and to do so in a way that evades pre-agreed notions of fairness. In contrast, the convivial society "would have the purpose of permitting all people to define the images of their own future" (Illich, 1975: 26). In the broad sweep of Illich's work, terms like 'education' have a special significance as the modulating output of industrialised society, in contradistinction to a more organic idea of 'learning' (Illich, 2000). For Illich, this decontaminated and refreshed idea of learning is core to the idea of a good life. We need to refocus from narrow ideas of machine learning to the broader idea of learning as an activity that, rather than inducing helplessness in the face of data manipulation at scale, grows people's confidence in themselves and in their capacity to solve problems. Inasmuch as machine learning can support this, it can become a convivial technology. In these new forms of convivial technology may also lie the forgotten non-market patterns that Illich saw as society's refuge from industrial self-destruction. Since counterfoil research "involves the public by showing that the demands for freedom of any group or alliance can be identified with the implicit interest of all" (Illich, 1975: 98), we can speculate that the route to sustainability may include a role for algorithmic processes in in the general production of solidarity.

References:

Abbott E (2015) *Flatland: A Romance of Many Dimensions, Revised Edition*. CreateSpace Independent Publishing Platform

Brugger P (2001) From haunted brain to haunted science: A cognitive neuroscience view of paranormal and pseudoscientific thought. *Hauntings and poltergeists: Multidisciplinary perspectives*, ed. J. Houran & R. Lange, pp 195–213.

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

Buterin V (2016) Decentralized Court | /r/ethereum. *reddit*. Available from: https://www.reddit.com/r/ethereum/comments/4gigyd/decentralized_court/ (accessed 28 April 2016).

Costello N (2015) Confessions of a former US Air Force drone technician. *Al Jazeera*, 19th November. Available from: http://www.aljazeera.com/indepth/features/2016/04/confessions-air-force-drone-technician-afghanistan-160406114636155.html (accessed 28 April 2016).

Duhigg C (2012) How Companies Learn Your Secrets. *The New York Times*, 16th February. Available from: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (accessed 28 April 2016).

Earle J and Kerr I (2013) Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy. *Stanford Law Review Online* 66: 65.

Electronic Frontier Foundation (2014) NSA Primary Sources. *Electronic Frontier Foundation*. Available from: https://www.eff.org/nsa-spying/nsadocs (accessed 1 May 2014).

Fallah-Adl H (1995) Atmospheric Correction, *University of Maryland Institute for Advanced Computer Studies*. Available from: http://www.umiacs.umd.edu/labs/GC/atmo/ (accessed 29 April 2016).

Ferran L (2014) Ex-NSA Chief: 'We Kill People Based on Metadata'. *ABC News Blogs*. Available from: http://abcnews.go.com/blogs/headlines/2014/05/ex-nsa-chief-we-kill-people-based-on-metadata/ (accessed 15 May 2014).

Freire P (2000) *Pedagogy of the oppressed*. Continuum International Publishing Group.

Goodfellow I, Bengio Y and Courville A (2016) Deep Learning. Available from: http://www.deeplearningbook.org.

Gorner J (2013) Chicago police use heat list as strategy to prevent violence. *Chicago Tribune*.

Available from: http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list (accessed 2 May 2014).

Grothoff C and Porup JM (2016) The NSA's SKYNET program may be killing thousands of innocent people. *Ars Technica UK*. Available from: http://arstechnica.co.uk/security/2016/02/the-nsas-skynet-program-may-be-killing-thousands-of-innocent-people/ (accessed 11 April 2016).

Haraway D (1988) Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14(3): 575–599.

Hastie T, Tibshirani R, Friedman J, et al. (2003) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 1st ed. 2001. Corr. 3rd printing edition. New York: Springer.

Illich I (1975) *Tools for conviviality*. Glasgow: Fontana.

Illich I (2000) *Deschooling Society*. New edition edition. Marion Boyars Publishers Ltd.

Kafka F (2010) *The Trial*. Hollywood, Fla.: Simon & Brown.

Kulstad M and Carlin L (2013) Leibniz's Philosophy of Mind. Winter 2013. In: Zalta EN (ed.), *The Stanford Encyclopedia of Philosophy*. Available from: http://plato.stanford.edu/archives/win2013/entries/leibniz-mind/ (accessed 28 April 2016).

Larson, J. et al., 2016. How We Analyzed the COMPAS Recidivism Algorithm. ProPublica. Available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (Accessed July 16, 2016).

Lyon D (2014) Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society* 1(2): 2053951714541861.

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

Mann S (2013) Veilance and reciprocal transparency: Surveillance versus sousveillance, AR glass, lifeglogging, and wearable computing. In: *2013 IEEE International Symposium on Technology and Society (ISTAS)*, pp. 1–12.

Mann S and Ferenbok J (2013) New Media and the Power Politics of Sousveillance in a Surveillance-Dominated World. *Surveillance & Society* 11(1/2): 18–34.

Massumi B (2005) The Future Birth of the Affective Fact. In: *Scribd*. Available from:

https://www.scribd.com/doc/242453979/Affective-Fact-Massumi (accessed 15 March 2016).

McCulloch, W.S. & Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), pp.115–133.

Metropolis N (1987) The beginning of the Monte Carlo method. *Los Alamos Science* (15.584): 125–130.

Mordvintsev, A., Olah, C. & Tyka, M., 2015. Inceptionism: Going Deeper into Neural Networks. *Google Research Blog*. Available at: http://googleresearch.blogspot.com/2015/06/inceptionism-going-deeper-into-neural.html (Accessed April 28, 2016).

Morton SG and Combe G (1839) *Crania americana; or, A comparative view of the skulls of various aboriginal nations of North and South America. To which is prefixed an essay on the varieties of the human species*. Philadelphia, J. Dobson; London, Simpkin, Marshall & co. Available from: http://archive.org/details/Craniaamericana00Mort (accessed 28 April 2016).

NYU School of Law (2016) Tyranny of the Algorithm? Predictive Analytics & Human Rights, *NYU Law School*. Available from: http://www.law.nyu.edu/bernstein-institute/conference-2016 (accessed 29 April 2016).

Piech C and Ng A (2012) K Means. *Artificial Intelligence Principles and Techniques; Stanford University*. Available from: http://stanford.edu/~cpiech/cs221/handouts/kmeans.html (accessed 28 April 2016).

Rehavi MM and Starr SB (2012) *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences*. SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. Available from: http://papers.ssrn.com/abstract=1985377 (accessed 5 April 2016).

Rinpoche S (1996) *The Tibetan Book of Living and Dying*. Reprint edition. Rider.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), p.386.

Schmidhuber J (2013) Handwriting Recognition with Fast Deep Neural Nets & LSTM Recurrent Nets & Deep learning. Available from: http://people.idsia.ch/~juergen/handwriting.html (accessed 29 April 2016).

Schuppli S (2013) Atmospheric Correction & The Politics of Remote Image Processing. In:

Rubinstein D, Golding J, and Fisher A (eds), *On the Verge of Photography: Imaging Beyond Representation*, Birmingham Article Press, pp. 16–32. Available from: https://www.academia.edu/6778193/Atmospheric_Correction (accessed 29 April 2016).

Scott J (1999) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New edition, Yale University Press.

The Intercept (2015a) SKYNET: Applying Advanced Cloud-based Behavior Analytics. *The Intercept*. Available from: https://theintercept.com/document/2015/05/08/skynet-applying-advanced-cloud-based-behavior-analytics/ (accessed 28 April 2016).

The Intercept (2015b) SKYNET: Courier Detection via Machine Learning. *The Intercept*. Available from: https://theintercept.com/document/2015/05/08/skynet-courier/ (accessed 28 April 2016).

The Magna Carta Project (2015) Clause 39 — Commentary for academic researchers, *The Magna Carta Project*. Available from: http://magnacarta.cmp.uea.ac.uk/read/magna_carta_1215/Clause_39/aca (accessed 30 April 2016).

transmediale (2015) *Matteo Pasquinelli - All Watched Over by Algorithms*. Available from: https://www.youtube.com/watch?v=So-miQplyd4 (accessed 28 April 2016).

Twachtman J (2013) Predicting the Epicenter of Crime: Analytics Tool Cuts Crime Rates, *Government Technology Magazine*. Available from: http://www.govtech.com/public-safety/Predicting-the-Epicenter-of-Crime-Analytics-Tool-Cuts-Crime-Rates.html (accessed 28 April 2016).

UNews (2015) Programming and prejudice | Utah computer scientists discover how to find bias in algorithms, *University of Utah*. Available from: http://unews.utah.edu/programming-and-prejudice/ (accessed 29 April 2016).

Wigner, E.P., 1960. The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Communications on Pure and Applied Mathematics*, 13(1), pp.1–14.

Yee S and Chu T (2015) A Visual Introduction to Machine Learning. Available from: http://www.r2d3.us/visual-intro-to-machine-learning-part-1/ (accessed 28 April 2016).