Objective Functions, Deep Learning and Random Forests

Alan F. Blackwell

Contribution to *Science in the Forest, Science in the Past*

Needham Institute, Cambridge. May 2017

Introduction: Science

A computer scientist seems an odd choice to speak either about science in the forest, or science in the past. Computer science is more often located in cities and offices than in forests, and is concerned with the challenges of the future rather than the past. ‘Science’ appears to be a point of enquiry shared with this symposium, but even this word is open to debate. It is often observed that a discipline including the word ‘science’ in its name introduces doubt as to why the claim is necessary. While Cambridge has long admired natural philosophy and the ‘natural sciences’, computer science fails to qualify as one of them. We computer scientists do not study nature, but only what we make ourselves[1]. A computer scientist is perhaps more akin to a novelist, sculptor or carpenter than to an astronomer or entomologist. This may be why computer scientists are unusually sensitive to the question of whether their work is objective.

In this paper, I discuss some critical contemporary questions for the discipline of computer science. Whether or not they relate to forest, the past, or even science, they do touch on the questions that Geoffrey Lloyd has set for us. My own research is concerned with the nature of knowledge, with the ways that we represent reality, and with the differing ways that words are used to reflect states of affairs as perceived by different academic communities. Even more centrally, I am concerned with what we achieve through making descriptions - with knowledge as a sociotechnical process rather than achievement.

Computer science provides an especially valuable perspective in this last respect, because of the ways that knowledge (as information, data) becomes a mathematical object subject to mechanical manipulation. Despite the fact that computer scientists prefer to align themselves with positivist traditions, through representational strategies in which binary 1s and 0s correspond to truth and falsity, much computer science research might be better characterised as a kind of mechanical sophistry, in which only the processing of information matters, and not any sense in which that

[1] Mahoney (1997) describes CS as “an amalgam of mathematical theory, engineering practice and craft skill”. Thanks to Willard McCarty for this reference.

information corresponds to the world outside the computer.

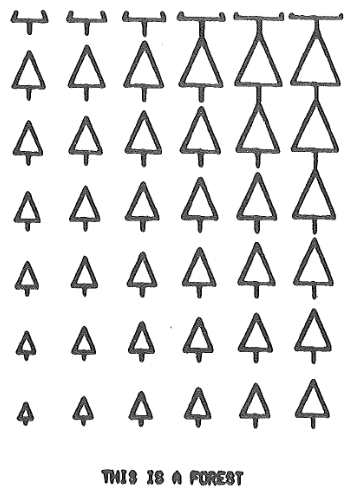
A phenomenology of the forest

Merleau-Ponty's phenomenology of space was a seminal characterization of embodied human experience within the built environment and with manufactured artefacts. In many ways it anticipated the work of my own specialist research field within computer science, which studies the human experience of information artefacts. However, I have always been nervously aware that the ordered descriptions that we achieve, in these sciences of the artificial, fail to capture a fundamentally different kind of embodied experience.

To a New Zealander, even one so poorly schooled in *Maoritanga* as myself, the English forest seems an anaemic affair. The New Zealand temperate rainforest that is known to *pakeha* (white New Zealanders) simply as 'the bush' is an archetypically blooming and buzzing confusion of thick-looping supplejack vines, viciously barbed bush lawyer, slick lichenous mud and ferns that scratch constantly around the knees. Especially in the mountain ranges of my native North Island - the Tararuas, impenetrable Ureweras, and the precipitous Waitakeres - marked trails are rare, and the shade is damp and gloomy under ancient trees that are easily imagined as ancestors or gods.

A New Zealander resident in England soon learns that the forests of his new home are purely geographic designations, not to be compared to the bush, or even the dense woods of European fairytale. The rather sparse trees to be seen at famous sites such as Nottingham Forest struggle to attain the stature they had assumed since one's childhood from popular legend.

To a New Zealander who works as a computer scientist and artificial intelligence engineer, the metaphor of the 'forest' in technical nomenclature departs even further from the numinous experience of the bush:



Meads (1972)

The metaphors of technical nomenclature are the object of my investigation today. I wish to explore the ways in which rich and numinous human experience must be reduced, if it is to be mechanised and quantified. All of us can appreciate the myth (if not the actuality) of the Forest - a place whose embodied wealth of form and substance defies easy characterisation, in which the essence of its imaginative presence rests in the many layers of corporeal texture that lie behind, beyond and beneath any surface on which the eye might settle.

In contrast, the mechanical geometries of modernity seem to present every kind of surface that the forest cannot. Rectilinear, smooth, precise, shiny, the hard science associated with laboratory instruments and rational manufacture is a means by which we are able to measure and order otherwise messy affairs.

I am particularly interested in the phenomena that we see discussed in the newspapers as ‘big data’, ‘machine learning’ and ‘artificial intelligence’ - each of these offering a metaphor that demands critical investigation in the terms that Geoffrey Lloyd asks us to consider as interlocking questions of (a) translatability/mutual intelligibility and (b) ontology/reality.

The technicalities of artificial intelligence research hold a degree of complexity, but not so much as to be unexplainable in this short presentation. However, to the extent that they can be explained, the reality of their intent when compared to the metaphorical implications of their names, might be as stark as the contrast between the modernist pretension of my English suburban garden and the New Zealand bush.

Artificial intelligence and decision trees

To start by dispatching the slightly facile metaphor of my title, the ‘random forest’ is a popular technique in the modern machine learning approach to artificial intelligence research. As with many such techniques, random forest algorithms address a wholly routine scientific problem. Close observation of any situation unearths many details that defy orderly description, including aspects that may appear indeterminate or contradictory. Such situations were problematic for ‘good old fashioned’ artificial intelligence (GOFAI) of the kind that I studied myself in the 1980s, because GOFAI assumes an orderly situation, of the kind where algorithmic deliberations might be specified using a ‘decision tree.’

If we imagine an artificially intelligent robot servant planning a course of action via a sequence of decisions, then each choice must be specified by enumerating the available alternatives. The series of choices can be arranged into an abstract tree – a diagrammatic structure in which a course of action is decided by starting from a ‘root’, choosing one from several alternative ‘branches’, continuing from branch to branch until reaching a ‘leaf’ that reflects the ideal action. In order for this leaf action to be an intelligent one, each branch along the way should have been chosen on the basis of statistical observation from previous experience, tabulating the likely chances of a successful outcome in each anticipated circumstance.

Unfortunately, although the principle appears straightforward, the available data seldom are. An action that works well in one situation might be disastrous in another, and even the billions of prior observations in corporate server farms seem not to exhaust the potential for new situations. The random forest is a pragmatic approach to the problem, which can be taken as a typical illustration of the shift from GOFAI to machine learning. Since there is no single decision tree that adequately describes even a small sample of real human experience, we construct our statistical ‘artificial intelligence’ in the form of a random forest, within which each tree describes local regularities, in the hope that one or another of these trees might be of some value when the moment arises. Computer memory is cheap, so the random forest, like many such algorithms, works by collecting data until it (hopefully) finds all plausible statistical combinations.

The random forest algorithm, although not especially fashionable or distinctive, can be taken to illustrate the essential character of modern ‘machine learning’ methods. They are techniques for generating simple choices in the presence of complex data, often through profligate use of increasingly inexpensive digital memory chips.

The simple answers obtained from machine learning algorithms are typically even simpler than the relatively sophisticated branch-based planning of a GOFAI decision tree. A more common requirement is simply to predict a single numerical output value, using a much larger amount of stored input data. A typical challenge might be to predict tomorrow’s stock price, based on a great deal of potentially relevant information (today’s price, past prices, the weather, articles in the financial press, company accounts, political announcements, twitter messages from company customers and so on). The input information might be enormously complex - but the creators of financial AI systems can afford plenty of memory to store it! The output appears very simple - just one number.

Statistical machine learning: Two kinds of regression

The process of predicting a single number, based on patterns that have been learned from previous data, is described as a ‘regression’ task. So how do we turn a ‘forest’ of complex data into a single tidy quantity? The statistical regression techniques used in machine learning, as with much of modern statistics, originate in the mathematical study of heredity associated with eugenics. Francis Galton first formulated the principle of ‘regression towards mediocrity’ as the observation that the children of exceptional parents are less likely to be exceptional themselves.

‘Regression’ continues to be a core principle of machine learning. The machine observes a series of data points, whether stock market prices, gene sequences or backgammon moves. Individual points might fit no obvious pattern, but over time, some kind of trend can be found, as the regression function tends back to an underlying average. This is only slightly over-simplified. If many kinds of data are observed, the trend relating them becomes much harder to visualise, because the lines must be drawn in a multi-dimensional space (one dimension for every variable). Nevertheless, the principle remains the same - that the ‘machine learning’ system predicts some quantity on the basis that it is the most mediocre explanation after all exceptions to the pattern have been discounted.

The second kind of regression is ‘logistic regression’, in which the machine learns to predict a categorical observation (yes/no, or animal/vegetable/mineral) rather than a numerical one. Once again, the term first appeared in the eugenics literature, being used to predict the likelihood over time that a parent will have a genetically defective child (Haldane & Smith 1947). The word ‘logistic’ is slightly problematic. As used by statisticians, the ‘logistic function’ is a logarithmic population curve, originally described by Belgian Pierre François Verhulst (1845) as a *fonction logistique*. While it seems that modern French uses this word *logistique* mainly in the sense of freight transportation logistics, Verhulst was referring to Napier’s logarithms.

Logarithm itself is an odd word, coined by Napier to describe his method of manipulating ratios. It seems that Napier himself did not explain his coinage, and the OED suggests alternative derivations related to arithmetic (*ars metrica*, the art of metrication), and also to logos, whose classical implications I must ask others at this symposium to explain. ‘Logistic’ might previously have described the ratio between two numbers, a rational process, or a rhetorical argument, in addition to its potential use as a mathematical term. So in modern usage, we are left with a logistic function that offers a logical approach to the logistics of formal logic. It may be derived from Napier’s exploration of ratio, associated with rationalizing, rationality and rational (ration-al) numbers. In the mathematical foundations of computer science, terminology seems constantly to drift between associations of quantification and of linguistic argumentation.

Whether described as linear or logistic regression, the epistemology underlying mysterious new developments such as ‘deep learning’ addresses issues already familiar from school statistics. We all remember vaguely how to reduce ‘big data’ to a single quantity by calculating the average (‘mediocre’), to estimate the trend line through a sequence of varying observations (‘linear regression’), or assess whether a repeatedly tossed coin is fair or not (‘logistic regression’). These simple intuitions start to falter only where there are many dimensions of data - the rich forest, rather than the two-dimensional chart or map. In many dimensions, there are many possible ways to draw a line between points, or to draw a boundary dividing fair coins from loaded ones. Each possibility must be assessed, in order to see which is the best of the many alternative explanations, each simple in itself. The process of assessment involves searching through the many alternatives, turning this way and that as each possible explanation seems marginally more likely than the last. The decisions to be made at each turning are of the same kind as the decision trees in the random forest that we started with.

And here we have another experience that is familiar to anyone who has been lost in the dense and mountainous forest of the New Zealand bush. In order to gain an idea of one’s location, it is necessary to climb out of a valley onto a ridge. Having attained the ridge (if the trees are not too dense), it is often possible to see the next ridge, but no further. It is very hard to tell, in a rich landscape, how the land lies, or where the highest peak might be. The many algorithmic techniques of machine learning are primarily concerned with this problem. Random forests, neural networks, genetic algorithms - despite their evocative names, all are simply strategies for finding the most effective simple explanations within a hugely varying and obscured set of possibilities, while avoiding the ‘local maximum’ of an explanation that accounts for some variations, but not others more distant.

This represents a theory of knowledge, in which a variety of possible explanations (‘models’) are tested against huge amounts of data to see which of them fits it best. The quality of the ‘fit’ is the central epistemological feature in this theory of knowledge. Fit quality must be expressed numerically, in order that one possibility can be tested against another in a convergence equation. This numerical function must be suitable for optimising search, guiding local decisions toward an unknown peak, without getting stuck among the trees of a local ridge. As with all search algorithms, the path to be taken emerges from the way that the goal or objective is described.

It is this ‘objective function’ that offers the epistemological meta-theory of all machine learning-based artificial intelligence. Once again, we see that the English language has introduced a convenient ambiguity, allowing us to make a semantic shift through double-reading of a key term. In the theory of search algorithms, the word ‘objective’ simply describes a goal, carrying no more metaphysical weight than the destination programmed into the satellite navigation system of your car. But when encountered within a theory of knowledge, the ‘objective function’ seems to assume an impressive status: of mechanical reasoning that will be unsullied by human subjectivity.

Ultimately, the procedure of ‘logistic’ regression, as guided by an ‘objective’ function, is not at all the objective and logical foundation of knowledge that these words might seem to imply. As I have shown, the terminology of machine learning has developed via engineering applications from the mathematics of eugenics and of robot navigation, and has only minimal relevance to the philosophical questions that theorists of artificial intelligence hope to address. It is not that mathematicians misunderstand the words they use. If pressed, any expert in the field would explain that the technical term ‘objective’ should not be taken to mean ‘objectivity’ in any sense, and that ‘logistic’ certainly does not imply ‘logical’. Unfortunately, such careful clarifications are seldom necessary among mathematicians themselves. Many computer scientists are poorly trained in basic principles of epistemology, while many philosophers are poorly trained in basic principles of engineering, meaning that they happily talk at cross-purposes with the aid of ambiguous terminology that neither properly understands.

Oracles and ground truth

At this point, we can look more closely at how the objective function is applied. Recall that the purpose of a statistical machine learning system is to find a model that best predicts or categorises the simple regularities within a large amount of varying data. In order to compare possible alternative models, the objective function may act as an ‘oracle’ (another technical term) judging which is the better. As with all prediction, this is far more easily done if we already know the answer.

This is precisely how artificial intelligence systems are trained - by showing them past cases in which we do know the answer, so that the objective function may compare each possible model to this ‘ground truth’ (the technical term is supervised learning). In my forest analogy, the ground truth is like a map that we had in our pocket all along, only pretending to be lost among the trees, while actually referring to our hidden map every time that we had to make a judgment on whether to walk to the next ridge.

Where might this ground truth come from, in these supervised machine learning systems? The sadly mundane answer is often that thousands of people are paid pennies to create a ‘ground truth’ by providing labels for large data sets of training examples. These workers are likely to be recruited via crowd-sourcing systems such as Amazon’s Mechanical Turk - a brazen reference to an earlier mechanical intelligence hoax. In this case, the ‘objective function’ is no more or less than a comparison of the trained model to previous answers given by the Mechanical Turkers. If the artificially intelligent computer appears to have duplicated human performance, in the terms anticipated by the celebrated Turing Test, the reason for this achievement is quite plain - the performance appears human because it *is* human!

The identity of the Turkers is kept secret, and this is the whole point (Irani & Silberman 2013). The people who collect the big data, store and process it in server farms, and replay the 'ground truth' of human interpretation have no desire to attribute their artificially intelligent creations to low-paid digital piece-workers. The irony of this situation is that the 'objective function' is impressively intelligent only to the extent that it replays human *subjective* judgments. The artificial intelligence industry is a subjectivity factory, appropriating human judgments, replaying them through machines, and then claiming epistemological authority by calling it logically 'objective' through a double reading of historically ambiguous mathematical terms.

Logistic regression and objective categorization

Logistic regression seems appealing in situations where machine judgment is necessary for the purposes of 'objective' classification. A typical question might be "is this person a criminal," to be answered objectively either yes or no. If we have information about all persons known to be convicted criminals - such as their bank balance, educational history, shoe size, head circumference, length of nose, colour of skin and so on - and corresponding information about people who are *not* criminals, then we can train a logistic regression system to tell us whether a given person on the street is, or is not, likely to be a criminal.

Some of these pieces of information (noses, skin etc) may be more or less useful in predicting which people are criminals or not, but this is not a problem if we have plenty of cheap computer memory. We can store all of this 'big data' information – ideally as much as possible – and leave the search algorithm to find which of it is useful and how it should be weighted. More of a problem is the ground truth on which the objective function is based. Are we confident that the ground truth, whether based on convictions or prison sentences, is wholly reliable? It is disappointingly easy to introduce circular reasoning when large data sets have been collected just in case they might be useful. Imagine if an arrest record were taken as part of the ground truth. It might easily be the case that being arrested is associated with having dark skin, in which case the objective function leads to the conclusion that a person with dark skin should be arrested. The old adage that correlation does not imply causation applies just as surely to the correlations underlying artificial intelligence systems. Embedding the correlation in an 'objective function' makes its causal interpretation no more objective!

Unfortunately, applications of artificial intelligence methods to police and security work are seldom published in the research literature (although, see Bennett Moses & Chan 2016), so it is hard to assess the degree of caution currently being applied, despite the rapid deployment of such systems around the world. More commonly, published research into classification might involve relatively innocuous logistic regression tasks. Is this a photograph of a horse? Is it a cat? Is a fish? A deep learning system considers huge numbers of features within an image that might represent fur or water, or anything else, and then correlates these features with previous images that it 'knows' (from the ground truth label assigned by a crowd sourcing worker, exploited graduate student or Facebook user) to belong to the relevant category.

Distinguishing photographs of fish from photographs of horses may be innocuous, but neither is it particularly impressive. Research sub-fields of artificial intelligence rely on classifications that are not too offensive, while still being sufficiently interpretive that the results do not seem trivial. One such is 'affective computing', in which images of human faces are classified according to whether they seem happy or sad. This can be challenging when the crowd-source workers are not certain whether a given person appears happy or sad, in which case the ground truth may be unreliable unless collected from the hammiest kinds of actor under instruction to express an unambiguous emotion. Indeed, many people being photographed (if not actors) find it difficult to tell us whether they are feeling happy or sad, in which case it is necessary to find an objective alternative – perhaps the amount of serotonin in a blood sample, or functional MRI scans that can tell us if our *body* or *brain* is happy, whether we know it or not.

The need for an objective function is a constant challenge of artificial intelligence research. I recently met a post-doctoral researcher who was training a classifier, using a form of logistic regression, to determine whether the person in a photograph was male or female. As usual, the training set consisted of photographs that had been labeled as either male or female by crowdsourcing workers. The researcher was pleased with results, which were achieving a high degree of 'accuracy' (which is to say, consistency with the prior labels). His suggestion was that this system might in future provide an independent and objective guide to gender.

Our conversation took place in a week when universities had been drawn into controversy in relation to the gender identity of transsexuals[2], which led me to worry how this system would respond if asked to classify photographs of a transgender person. The researcher, who was visiting from a country in which gender identity is less openly debated than in the UK, considered that this would not be a problem, because his system would report the 'real' gender of the person. To a technologist having little interest in the world of gender politics, the engineering problem seemed to involve straightforward binary classification, in which the objective truth could be determined from the body far more reliably than one's emotional state (perhaps by removing clothes, or if necessary, inspecting chromosomes).

[2] See Steven Morris, writing in *The Guardian*, Wednesday 18 November 2015. "Germaine Greer gives university lecture despite campaign to silence her" https://www.theguardian.com/books/2015/nov/18/transgender-activists-protest-germaine-greer-lecture-cardiff-university

Objective functions and text

The discussion of statistical machine learning so far has attempted to demystify the prediction of stock markets, and 'recognition' of photographic images, emotions, boys and girls, criminals and so on. Many other headline examples celebrating the achievements of artificial intelligence are similarly straightforward. Board games such as backgammon, chess or go, are equally trivial, in the sense that they all have a clearly defined objective function. Although humans may find some of these tasks difficult, they are all quite clearly mechanical (which is to say, unlike forests in being observable, differentiable, shiny, rectilinear and so on), and thus easily amenable to mathematical description.

However, some of the greatest challenges and apparent achievements for artificial intelligence are related to the processing of text, rather than numbers. How can an objective function be achieved when constructing a text?

Consider the following snippets of conversation between a human on one side, and an advanced artificial intelligence on the other:

> Human question: Why does Juliet die?
>
> Computer answer: *Juliet sees Romeo dead beside her, and surmises from the empty vial that he has drunk poison.*
>
> Human question: Tell me what Donald Trump will do next?
>
> Computer answer: *Donald Trump looks set to be a controversial and unpredictable President after an inflammatory election campaign.*

These exchanges would be quite unremarkable between two humans. But between a human and a computer, they do seem remarkable. This conversation seems to pass the 'Turing Test' – the imitation game in which a computer converses in a manner indistinguishable from a human. However, the exchanges above did not surprise me at the time … because this is the transcription of an actual 'conversation' in which I was the human, and the computer was simply my web browser, which passed the questions I had typed to Google. The 'answers' were text returned in my browser window by the Google search engine.

When viewed in this light, the conversation is hardly impressive at all. Fifty years ago, these exchanges might have seemed magical, evidence that artificial intelligence has been achieved. Today, they are so far from magical as to induce a yawn. This is not intelligence, it is simply another Google search – an everyday transaction. Yet the mechanisms by which this once-intelligent, now-mundane interaction have been achieved are the same kinds of statistical process already described. Google algorithms work by calculating statistical correspondences between the words that I type, and web pages that might provide pertinent information. The sentence that is returned appears to have been written by a human, because indeed it has been written by a human – the author of another web page.

I should note, in passing, that the things I actually saw on my computer screen during this exchange contained many clues that would alert an observer from 50 years ago to the relative absence of intelligence. The answers as reported in my transcript above *appear* intelligent as I quoted them, because of the fact that I transcribed with a degree of interpretation – for example, ignoring the 'Google' logo that appears at the top of the screen, knowing where to type, judiciously ignoring the text of an advertisement that is 'clearly' (to me) not relevant to my question and so on. I am able to make all of these interpretive judgments routinely and unconsciously, precisely because Google searches are so familiar to me. For the 1970 observer, my own interpretation and transcription of the user interface might seem as foreign (or even more foreign) than the responses provided by the search engine. This fundamental property of interaction with machines is described by Collins and Kusch (1999) as *Repair, Attribution and all That* (RAT) – human users constantly 'repair' the inadequacy of computer behavior, then attribute the results to intelligence on the part of the machine, while discounting the actual intelligence that was supplied in the process of repair.

So to return to my transcribed 'conversation', now that we know the 'answers' are simply text copied from a Google search result, it is quite obvious that these texts were written by human authors. For the first question, we might surmise that the words were written by the author of a school study guide, and for the second, by a journalist or political commentator. If an *employee* of Google were to pretend that he or she was the actual author of these pages, then this pretense would be (morally) plagiarism and (commercially) an infringement of copyright. On the other hand, if we pretend that the Google *algorithm* was the author, as I have done, then would this algorithmic 'artificial intelligence' be guilty of plagiarism or copyright infringement? Can an algorithm be guilty of anything? It is not morally culpable, and cannot be tried in a court of law, despite the fact that using the phrase 'artificial intelligence' as a noun implies some kind of legal personhood.

In the case of the relatively simple and honest behaviour of the Google search engine, the question does not arise. Google is quite clear in stating (and ample legal precedent has confirmed) that it is simply providing an indexing service, not claiming to be the author of the content it delivers. (This despite the fact that the text I quote above was copied directly from the Google search results page, never visiting the sites created by the actual authors, with the result that the index has to some extent *become* the text through my reading of it).

The situation is more complex when an algorithm 'mashes up' text that was written by multiple authors rather than a single identifiable person. It is quite routine to create algorithms that generate text, on the basis that after a short sequence of words, it is statistically straightforward to predict the next word. This behaviour is seen every day, in our search bars and web browsers. If I type "How may I …", Google helpfully offers to complete my sentence: "How may I assist you", on the basis of statistical likelihood. Presumably some actual person (more likely many) has previously typed these words so that Google may respond with this 'intelligent' assistance. It would be silly to suggest that Google has infringed their individual copyrights, or plagiarised the work of a multitude, who themselves have only repeated a cliché.

On the other hand, if I type "It is a tr…", Google immediately offers "It is a truth universally acknowledged". We know that Jane Austen is the author of this phrase, but this is not attributed or acknowledged. Undoubtedly many people have typed these words, but it is Google that offers them to me, with no mention of the original author. It might be suggested that Google is providing me with an 'intelligent' plagiarist, to be defended on the convenient statistical principle that we have only engaged in a statistical transaction, with no intension that the result should be passed off as the work of a creative novelist.

Summary

We are often told that in an imminent era of automated 'general intelligence,' computers will acquire creative capabilities, acting on their own initiative, and perhaps even presenting a threat to the future of the human race, as they decide autonomously to act in their own interests rather than ours. Yet we have seen that the 'intelligent' behaviour of machines is no more than human behaviour reflected back to us. The ground of mutual intelligibility, between the artificially intelligent ontologies of the machine world, and our embodied experience of the human one, may not be as hard to discover as one would suppose.

On the contrary, one might argue that a computer using 'deep learning' techniques to produce text or music is no more displaying intelligence than a television is interpreting Beethoven during a broadcast from the Albert Hall. The orchestra still exists – it is simply playing elsewhere. The machine is only a medium, not an interpreter. It can take us time to recover from the surprising form of new technologies, and to recognise the ways in which they yet again reconfigure social relations. Nevertheless, this is all that is happening. Even when associated with whole forests of data, the objective ground truth can usually be traced to subjective judgments, made by a person or persons who are more or less hidden from view.

I introduced this paper with the suggestion that interlocking questions of (a) translatability/mutual intelligibility and (b) ontology/reality might be explored, in relation to the phenomena of 'big data', 'machine learning' and 'artificial intelligence'. Although my titular reference to random forest algorithms was a somewhat facile connection to the theme of this symposium, it offered a starting point as good as any, in introducing the problems of situated and embodied complexity versus mechanical objectivity that were my key critical concerns. If we are going to consider artificial intelligence as a matter of philosophical interest, then we must pay more attention to the actual algorithmic basis through which reality is represented, ontologies are constructed, and human observers interpret the interactions that they experience with complex systems.

I have also discussed the way that words pass among disciplines, acquiring new connotations that unfortunately represent wishful thinking on the part of researchers, and over-excitement among critics, rather than rigorous analysis. The 'objective function' is perhaps the most egregious case, sliding from a purely mathematical optimization principle to an anachronistically positivist interpretation of machine learning. Common narratives of general artificial intelligence, derived more from science fiction than from cognitive science, all too often serve to dehumanize the real people whose work is hidden from view, and to empower and enrich the corporations whose wealth is derived from that work.

If we persist in personifying software systems as heroic cyborg servant-champions (AlphaGo, Watson, Alexa, Siri), whether benevolent or malign, we run the risk of indemnifying those actually responsible for their design and operation. A software system is not legally accountable, so if a crime is committed, we open ourselves to the ludicrous defense that nobody did it. Instead we must recognize what the phrase 'Big Data' hides in plain sight – that the Internet is an astonishing bureaucracy, thousands of times more massive than any created before, and intruding on every rich aspect of human life from friendship to family. Meanwhile, each logical step taken by an artificial intelligence system has been extracted, averaged, and abstracted from the diverse multitudes, even forests, of real life. A contextualized, qualitative and interpretive social science is needed to defend the true complexity and diversity of embodied and situated human experience.

References

Bennett Moses, L., & Chan, J. (2016). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society*, 1-17.

Blackwell, A.F. (2015). Interacting with an inferred world: The challenge of machine learning for *humane* computer interaction. In *Proceedings of Critical Alternatives: The 5th Decennial Aarhus Conference*, pp. 169-180.

Collins, H., and Kusch, M. (1999). *The shape of actions: What humans and machines can do*. MIT press.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2), 215-242.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246-263.

Haldane, J. B. S., and Smith, C. A. (1947). A simple exact test for birth-order effect. *Annals of Eugenics*, *14*(1), 117-124.

Irani, L. C., & Silberman, M. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 611-620). ACM.

Mahoney, M. S. (1997). Computer Science: The Search for a Mathematical Theory. In J. Krige and D. Pestre (eds). *Science in the Twentieth Century,* Amsterdam: Harwood Academic Publishers, pp. 617-34.

Meads, J.A. (1972) A terminal control system. In F. Nake and A Rosenfeld (eds), *Proc. IFIP Working Conference on Graphic Languages*, pp. 271-287

Merleau-Ponty, M. tr: Colin Smith (1945/1965). *Phenomenology of Perception*. London: Routledge & Kegan Paul.

Verhulst, P. F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux mémoires de l'académie royale des sciences et belles-lettres de Bruxelles*, *18*, 14-54.