Simulate, optimise, partition: algorithmic diagrams of pattern recognition from 1953 onwards

Adrian Mackenzie

Sociology, Lancaster University

a.mackenzie@lancaster.ac.uk

Contemporary attempts to find patterns in data, ranging from the now mundane technologies of touchscreen gesture recognition through to mammoth infrastructure-heavy practices of deep learning conducted by major business, scientific and government actors to find cats (Markoff 2012), the Higgs boson, credit card fraud or terrorists, rely on group of algorithms intensively developed during the 1950-60s in physics, engineering and psychology. Whether we designate them as pattern recognition, data mining or machine learning (all terms that first came into play during the 1950s), the standard account enunciated by proponents (and opponents) of these techniques is that they uncover patterns in data that cannot appear directly to the human eye, either because there are too many items for anyone to look at, or because the patterns are too subtly woven through in the data.

In the contemporary narratives of their efficacy and indeed necessity, the spectrum of differences accommodated under the rubric of pattern is striking. Pattern here is understood to encompass language, images, measurements and traces of many different kinds. Pattern finding techniques – although that term is problematic because it suggests skilled hands doing something; I will refer to them as *operations* – diagram a strikingly new kind of continuum or field which accommodates seemingly very different things – terrorists, fundamental particles, photographs, market transactions, utterances and gestures – more or less uniformly.

What counts as pattern finding today, I will suggest, can be better understood by taking into account the transformations in simulating, optimising and above all *classifying* associated with different uses of computers taking shape in the mid-twentieth century. Despite their often somewhat a-historical invocations, the patterns recognised in pattern recognition have an historically concrete specificity. From the plethora of operations in current use, three diagrams developed in the Cold War era operate in contemporary modes of pattern finding:

1. Monte Carlo simulation as a way of shaping flows of random numbers to explore *irregular probability distributions;*
2. *convex optimisations* or finding maximum or minimum value numerical solutions to systems of equations as a way of classifying things;
3. *recursive partitioning* algorithms that re-order differences according to clustering and sparsity in data.

The operations expressed in these diagrams took shape in different places – nuclear physics, control systems engineering and psychology, but soon moved across boundaries between academic disciplines, and between domains such as universities, industry, the military, and government. Each of them, deeply reliant on electronic computation, configures a different mode of moving through data in order to find or make patterns. The different perspectives on event, difference, and recognition they embody imbue many power relations, forms of value and the play of truth/falsehood today. In each case, they contributed something to the formation of an operational field that today has generalized to include many aspects of media, culture, science, business and government, none of which exists purely or in isolation, but in combination with each other. Because the diagrammatic operations of probability, optimisation and classification have intersected, we today increasingly inhabit a pattern-recognised space that configures what it is to belong, to participate, to anticipate, to speak or to decide differently.

What are continuities in the operational field?

> If the operatives of the Cold War could reserve for themselves the position of gray eminence, the distant advisor to the executive power, the new spaces of collectively intelligent networks and the asymmetrical relations these put in place demand instead the more difficult position of gray immanence (Fuller and Goffey 2012, 32).

In order to understand how the generalization of probability-simulation-optimization took place, Matthew Fuller and Andrew Goffey's contrast between gray eminence and gray immanence is suggestive. The shift between grey eminence or powerful technical-scientific advisors to executive power and grey immanence in intelligent networks is precisely the movement that we might delineate by paying attention to the operations of simulation, optimisation and partitioning that underpin and in many ways sub-structure social network media, contemporary health and biomedical knowledges and credit ratings, to name a few. The operations for working with data, numbers, probability and categories took shape deep in the epistemic cultures of the Cold War. Specific locations such as the RAND Corporation, the Cornell Aeronautical Laboratory and IBM Corporation figure large here, but they somehow proliferate and disseminate in a field that is neither that of lifeworld (lived, urban, etc.) experience linked to a subject position (the grey eminences of Cold War science, as described for instance in *How Reason Almost Lost Its Mind* (Erickson et al. 2013) nor an unliveable mathematical abstraction (the Euclidean space of geometrical abstraction analysed by Cold War philosophers such as Hannah Arendt(Arendt 1998)), but something more like an *operational field* in Michel Foucault's sense of the term (Foucault 1972, 106).[1] The dimensions

---

[1]Foucault writes:

and composition of this field undergo constant expansion and accumulation, partitioning and aggregation via operations increasingly detached from expert 'gray eminences.' Through it, in it, near it, derived from it, many different artifices, devices, arrangements, operations and processes accumulate. It is an power-laden operational space that traverses and structures many aspects of our lives, but is only intermittently sensible or palpable to us. It appears at the intersection of many scientific disciplines, various infrastructures, operations, and institutions. It is object and domain of much work and investment in management, enterprise and State.

Authoritative and recent accounts of Cold War rationality have strongly dramatised the part that cybernetics played in re-structuring human-machine-organism relations in various Cold War contexts ranging across engineering, psychology, management, military strategy and anthropology (Bowker 1993; Edwards 1996; Hayles 1999; Pickering 2009; Halpern 2015). These accounts argue that cybernetic systems of control, automation and cognition were constitutively metaphorical and openly trans-disciplinary. From the outset, cybernetics propositions concerned organisms, organisations, states, machines and subjectivity. The diagrammatic operations I describe here are also intimately linked with transformations in what counts as pattern, recognition, learning and intelligence, but their mode of semiosis somewhat differs from cybernetics. The composition of the operational field we are concerned with here lacks the constitutive generality of cybernetics. Although coeval with cybernetics, it is much more densely woven through methods, operations, infrastructures, forms of expertise, and models. Hence, we can't understand the contemporary force of the operational field without moving backwards and seeing how those power-laden operations proliferated and coalesced. It is an operational generalization rather than an enunciative one, and takes place, unsurprisingly given its contemporary materialisation, in code.
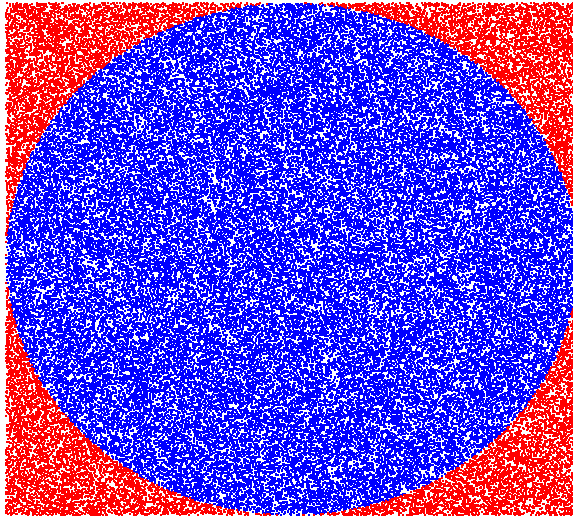
Figure 1: A Monte Carlo simulation of π

Exact means simulated, simulated means open

In 1953, Nicholas Metropolis, the Rosenbluths, and the Tellers, all physicists working at Los Alamos, were considering 'the properties of any substance which may be considered as composed of interacting individual molecules' (Metropolis et al. 1953, 1087). These properties might be, for instance, the flux of neutrons in a hydrogen bomb detonation. In their short, evocatively titled and still widely cited paper 'Equations of state for fast calculating machines' (over 20,000 citations according to Google Scholar; over 14,000 according to Thomson Reuters Web of Knowledge), they describe how they used computer simulation to manage with the inordinate number of possible interactions in a substance, and to thereby come up with a statistical description of the properties of the substance. While statistical descriptions of the properties of things are not new,[2] their model system consists of a square containing only a few hundred particles. (This space is a typical multivariate joint distribution.) These particles are at various distances from each other and exert forces (electric, magnetic, etc.) on each other dependent on the distance. In order to estimate the probability that the substance will be in any particular state (fissioning, vibrating, crystallising, cooling down, etc.), they needed to integrate over the many dimensional space comprising all the distance and forces between the particles. The dimensions of the space in which all of the variables describing the velocity, momentum, rotation, mass for each of the several hundred particles is already an expansive one. As they write, 'it is evidently impossible to carry out a several hundred dimensional integral by the usual numerical methods, so we resort to the Monte Carlo method' (1088), a method that Nicholas Metropolis and Stanislaw Ulam had several years previously already described in an earlier paper (Metropolis and Ulam 1949). Here the problem is that the turbulent randomness of events in a square containing a few hundred particles thwarts calculations of the physical properties of the substance. They substitute for that non-integrable turbulent randomness a controlled flow of random variables generated by a computer.

> I now realize that I could not define the statement as a unit of a linguistic type (superior to the phenomenon of the word, inferior to the text); but that I was dealing with an enunciative function that involved various units (these may sometimes be sentences, sometimes propositions; but they are sometimes made up of fragments of sentences , series or tables of signs, a set of propositions or equivalent formulations); and, instead of giving a 'meaning' to these units, this function relates them to a field of objects; instead of providing them with a subject, it opens up for them a number of possible subjective positions; instead of fixing their limits, it places them in a domain of coordination and coexistence; instead of determining their identity, it places them in a space in which they are used and repeated. In short, what has been discovered is not the atomic statement - with its apparent meaning, its origin, its limits, and its individuality - but the operational field of the enunciative function and the conditions according to which it reveals various units (which may be, but need not be, of a grammatical or logical order) (Foucault 1972, 106)

[2]Isabelle Stengers provides an excellent account of some of the nineteenth century development of thermodynamics (Stengers 2011). The history of statistics since the late seventeenth century obviously forms part of the background here [(Stigler 1986; Hacking 1975).

While still somewhat random (i.e. pseudo-random), these Monte Carlo variables taken together approximate to the integral, the area or volume under the curve geometrically understood, of the many dimensional space.

A toy example to show the intuition of Monte Carlo simulation is shown in Figure 1. The point of this simulation, which comprises a half dozen lines of code, is to calculate the value of the mathematical constant π, a value that describes the ratio between the radius and circumference of a circle. In this Monte Carlo simulation of π, 100000 points are randomly generated, each point described by an x-y coordinate. Given the formula for the area of a circle (πr^2) and assuming the radius of the circle is 1 unit, the algorithm tests each random point to see if falls inside the circle. The ratio for π is given by dividing the number of points inside the circle by the total number of randomly generated points and then multiplying by 4 (since if the radius of the circle =1, then the diameter =2, and therefore, the total area of the bounding box $= 2 \times 2$, so $\pi = 4 \times p$ the proportion inside the circle). The point of this demonstration is not to re-state the value of π, but to suggest that we can see here an inexact, probabilistic calculation of a number that previously epitomised mathematical precision and geometric ideal form (the circle). Monte Carlo simulation we might say puts ideal form on a computational footing.

Amongst the many different implications of this simulation of ideal forms, perhaps the most important concerns the status of probability. The contour plot in Figure 2 was generated by a variant of Monte Carlo simulation called MCMC – Markov Chain Monte Carlo simulation – that has greatly transformed much statistical practices since the early 1990s (see (Mcgrayne 2011) for a popular account). Like the simulation of π, this simulation calculates significant numbers, this time μ_1 and μ_2, the mean values of two probability distributions. This seemingly very simple simulation of the contours of two normally-distributed sets of numbers shows four main diagrammatic operations. First, pluri-dimensional fields arise at the intersection of different axes or vectors of variation. While the graphic plot here is two dimensional in this case, in can be generalized to much higher dimensions, thus combining many more variables. Second, although Descartes may have first formalised the coordinate geometry using axes in the form of the Cartesian plane, we can see in Figure 2 that this plane has a different consistency or texture. While the Cartesian axes are intact with their scales and marked intervals, the field itself is topographically shaped by a plural flux of random numbers in the Monte Carlo simulation. The topographic convention of showing heights using contour lines work in Figure 2 to render visible continuously varying distributions of values across multi-dimensions. The curves of these contours outline the distribution of values in large populations of random numbers generated in the simulation. While the curves of the contour lines join elevation that have the same value, the continuous undulation of values overflows line or regular geometrical form. Third, superimposed on the contour lines, a series of steps or a path explore the irregular topography of the pluri-dimensional field. This exploration appears in the dense mass of black points on the figure deriving from the further flows of random numbers
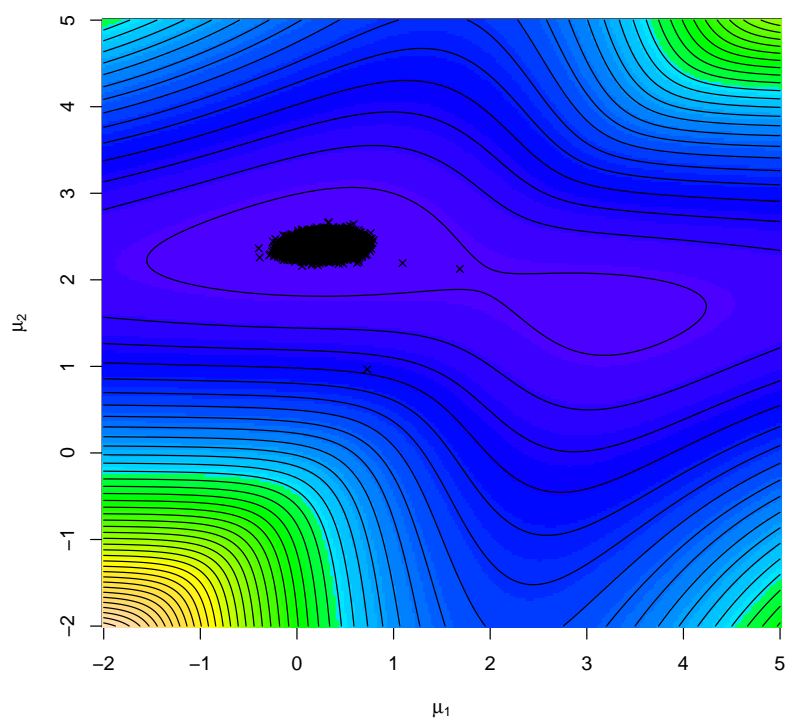
Figure 2: A Markov Chain Monte Carlo simulation of two normal distributions

moving towards the highest point, the peak of the field, guided by continuous testing of convergence. Finally, the plot as a whole graphs the different values of the means (μ_1, μ_2) of two random variables distributed over a range of different possible values. We need know nothing about what such variables relate to – they may refer to attributes of individuals, behaviours of markets, growth of an epidemic, the likelihood of an asthma attack. But as descriptions of probability distributions, as ways of assigning numbers to events, the MCMC simulations widely used today to explore complex topographies of things in variation suggest that this form of computation transforms pattern into a probabilistic simulation.

Optimise in order to learn: 1957

We know from the histories of Cold War rationality that cognition and calculation are tightly entwined. Cold War cognition calculates its chances of winning, error, loss, costs, times and delays in for instance game theory (Erickson et al. 2013, ch5). But the mechanisms and devices of this entwining of cognition and calculation are not necessarily obvious or homogeneous. The subtle mechanisms and diagrammatic operations of cognitive calculation sometimes remain opaque and almost subliminal in the words of Cold War discourse. Yet it is precisely these mechanisms that flow along long fault-lines into contemporary knowledge apparatuses with all their power-generating armatures in areas such as security or online media. The operations of these mechanisms are often quite localised and in some cases trivial (e.g. fitting a straight line to some points), but their operations accumulate and generalise in ways that sometimes produce strategic, even hegemonic effects in contemporary culture. While there are quite a few operations that might be examined from this perspective, the case of the Perceptron from 1957 is evocative because of both its cybernetic provenance and its subsequent re-enactments during the 1980-1990s in the form of neural nets, in the very abundant support vector machines of the 1990s (Cortes and Vapnik 1995), and today in the massively ramified form of 'deep learning' (Hinton and Salakhutdinov 2006).

'Learning' is pivotal to diagrammatic operation of machines such as the Perceptron and its many contemporary avatars (see (Mackenzie 2015a)). While initially framed in terms of rational actors playing games (for instance in 1944 in Oskar Morgenstern and John von Neumann's *Theory of Games and Behaviour* (Von Neumann and Morgenstern 2007); see (Erickson et al. 2013, 133-134)), the locus of learning shifts diagonally through diagrams, plans and equations into different arrangements typified by the Perceptron. Such devices stand at some remove from the agential dilemmas of Cold War military strategy that animated high profile game theory. While the Perceptron retains some figurative aspects of its biological inspiration in the neurone, these figurative aspects are rapidly overlaid and displaced by asignifying processes that have continued to mutate in subsequent decades. The so-called 'learning problem' and the subsequent theory of learning machines was developed largely by researchers in the

1960-1980s, but based on work already done in the 1950s on learning machines such as the Perceptron, the neural network model developed by the psychologist Frank Rosenblatt in the 1950s (Rosenblatt 1958). Drawing on McCulloch-Pitts model of the neurone, Rosenblatt implemented the Perceptron, which today would be called a single-layer neural network on a computer at the Cornell University Aeronautical Laboratory in 1957. A psychologist working in an aeronautical laboratory sounds rather odd, but given that the title of Rosenblatt's 1958 publication – 'The Perceptron: A probabilistic model for information storage and organization in the brain' – already suggested an intersection between statistics (probabilistic models), computation (information storage and organization), and neuroscience ('brain'), perhaps Rosenblatt's cross-campus mobility is symptomatic of the diagonal movement occurring around learning. The very term 'Perceptron' already amalgamates the organic and the organic, the psychological and the technological in a composite form.

Like the Monte Carlo simulations, the Perceptron operates according to a simple computational process: a machine can 'learn' by classifying things according to their position in a pluri-dimensional data space. 'Geometrically speaking,' writes Vladmir Vapnik (a machine learning researcher famous for his work on the support vector machine), 'the Perceptron divides the space X into two parts separated by a piecewise linear surface. … Learning in this model means finding appropriate coefficients for all neurons using given training data' (Vapnik 1999, 3). Note the emphasis on *classification:* If the Monte Carlo simulation generated a space in which many different variables could be integrated in exploring irregular probability distributions, devices such as the Perceptron configure a space in which the divisive operation of classification can be configured as a problem of optimisation. 'Learning' as Vapnik puts it, 'means finding appropriate coefficients.'

The practice of learning here owes more to logistics than perhaps to cognition or neuroscience. That is, learning occurs through and takes the form of optimisation. Optimisation in turn is understood in in terms mathematical functions located in high-dimensional spaces that cannot be analysed in closed-forms, but only explored looking for maxima or minima. Optimisation algorithms such as gradient descent or expectation maximisation (EM) are the key components here. That is, the theory of machine learning alongside decision theory was interwoven with a set of concepts, techniques and language drawn from statistics. Just as humans, crops, habitats, particles and economies had been previously, learning machines became entwined with statistical methods. Not only in their reliance on the linear model as a common starting point, but in theories of machine learning, statistical terms such as bias, error, likelihood and indeed an increasingly thorough-going probabilistic framing of learning machines emerged.

Learning machines optimise rather than cognize. The plot of a few points in a two dimensional space shown in Figure 3 again has to stand in for a much more voluminous, densely populated and pluri-dimensional space. The different shapes of the points index different categories of things (for example, `male`
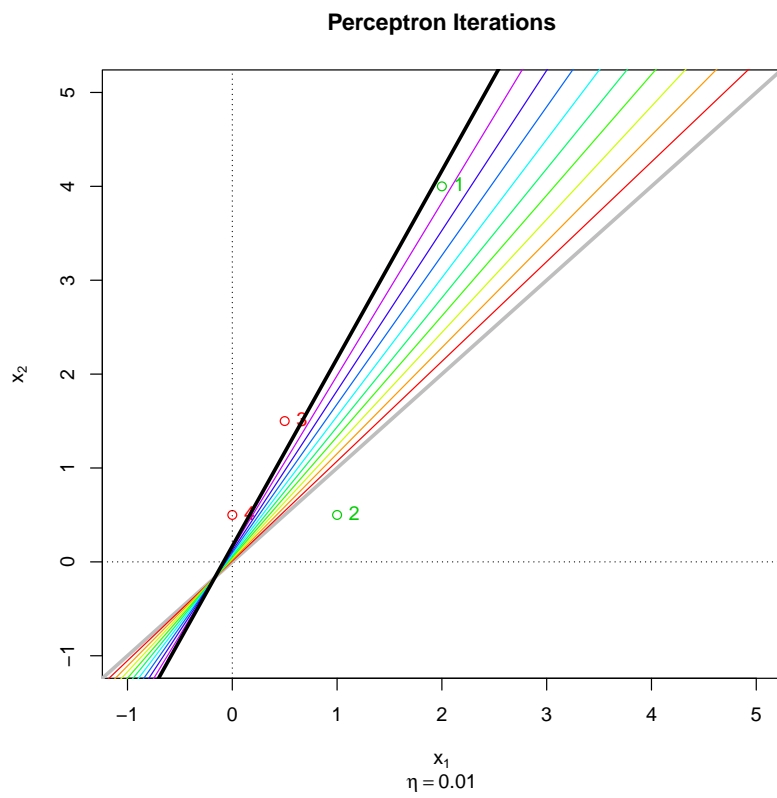
Figure 3: Perceptron learns to separate

vs. `female`). The lines in this figure are the work of a Perceptron learning to classifying the points by searching for lines that divide the space. Starting with an arbitrary line, the Perceptron tests whether a line effectively separates the different categories. If it does not cleanly separate them, the algorithm incrementally adjusts the parameters that define the slope and intercept until the line does run cleanly between the different categories. It easily may be that the different categories overlap, in which case, the Perceptron algorithm will never converge since it cannot find any line that separates them.

For our purposes the point is that any learning that occurs here lies quite a long way away from biological figure of the neurone. Some biological language remains – 'activation functions', for instance, figure in the code that produces the lines in Figure 3 – but the algorithmic process of testing lines and adjusting weights in order to find a line or plane or hyperplane (in higher dimensional data) very definitely rely on an optimisation process in which errors are gradually reduced to a minimum. Furthermore, the diagrammatic operation of the Perceptron – repeating drawing of lines in order to classify – appears in a number of variations in the following decades. Some of these variations – linear discriminant analysis, logistic regression, support vector machine – generalise the process of finding separating lines or planes in data in quite complicated ways in order to find more supple or flexible classifications. The Perceptron is not unique in doing this. Subsequent developments, including various machine learning models such as logistic regression, neural nets, support vector machines, k nearest neighbours and others present variations on the same diagrammatic operation: pattern recognition entails learning to classify; learning to classify means finding a way of best separating or partitioning a pluri-dimensional data space; the best partition optimises by reducing the number of misclassification errors. It is not surprising that later iterations and variations of the Perceptron proliferated the process of optimisation. For instance, neural nets, current favourite machine learning techniques for working with large archives of image and sound, aggregate many Perceptrons in layered networks, so that the Perceptron's diagrammatic operation of classification can be generalised to highly complex patterns and shapes in the field of data. No matter how complex the classifiers become, they still propagate the same diagrammatic operation of drawing a line and testing its power to separate.

Learning to classify in this way generated many new possibilities. Vapnik observes: 'the Perceptron was constructed to solve pattern recognition problems; in the simplest case this is the problem of constructing a rule for separating data of two different categories using given examples' (Vapnik 1999, 2). While computer scientists in artificial intelligence of the time, such as Marvin Minsky and Seymour Papert, were sceptical about the capacity of the Perceptron model to distinguish or 'learn' different patterns (Minsky and Papert 1969), later work showed that Perceptrons could 'learn universally.' For present purposes, the key point is not that neural networks, the present day incarnations of the Perceptron have turned out by the 1980s to be extremely powerful algorithms in learning to distinguish patterns, and that intense research in neural networks has led

to their ongoing development and increasing sophistication in many 'real world' applications (for instance, in commercial applications such as drug prediction (Dahl 2013)). Rather, the important point is that it began to configure computational machinery as an ongoing learning project.Trying to understand what machines can learn, and to predict how they will classify or predict became central concerns precisely because machines didn't seem to classify or predict things at all well. The project of learning to classify by optimising choice of coefficients or model parameters has become powerfully performative today, since it underpins many of the recommendation systems, the personalized or targetted advertising, and increasingly the shaping of flows of media, financial and security power. In all of these settings, classification has become a matter of learning to construct a rule for separating things.

Ramified decisions

Cold War information theory says information is that which allows a move down a decision tree (Solovey and Cravens 2012, 103). Decisions – as the term itself suggests – are a kind of cut. But decisions often have a complicated branching structure, at least, in the procedural rationality typical of Cold War operations research and logistics. In procedural rationality, branches in a decision tree stem from rules that embody the expert knowledge of the grey eminences of management science and its cognate power-knowledge formations. Increasingly during the decade of the Cold War, these rules were shaped by optimisation and modelling procedures that sought to allocate resources most efficiently (especially in the field of operations research; see (Erickson et al. 2013, 79)).

The decision tree was and remains a key diagrammatic construct in Cold War closed-world thinking in its attempt to represent optimal allocation of resources amidst systems of increasing scale and complexity. The development of procedural rationality based on various algorithmic procedures (the linear programming and dynamic programming techniques are core of much operations research (Bellman 1961)) was shadowed by the growth of a different form of decision tree: the classification and regression tree (Breiman et al. 1984). During the 1960s, the decision tree itself was computationally re-generated as rather different kind of device that in some ways owes more to older systems of classification associated with taxonomy or natural history. Expert decision rules are replaced by learning algorithms that partition data according to quasi-statistical measures of mixedness or purity. This inversion of the decision tree again permits its generalization. In a certain sense, the decision tree (and its contemporary incarnation in the very popular random forest (Breiman 2001)) dismantles the classical tree with its reference to kinds of being. It also obviates in certain ways the decision tree of procedural rationality as a distillation of expert knowledge. The decision tree is no longer a way of regulating information flow towards optimum resource allocation (missiles, cargoes, troops, etc.). In classification and regression trees, branches are instead something to be learned from the data
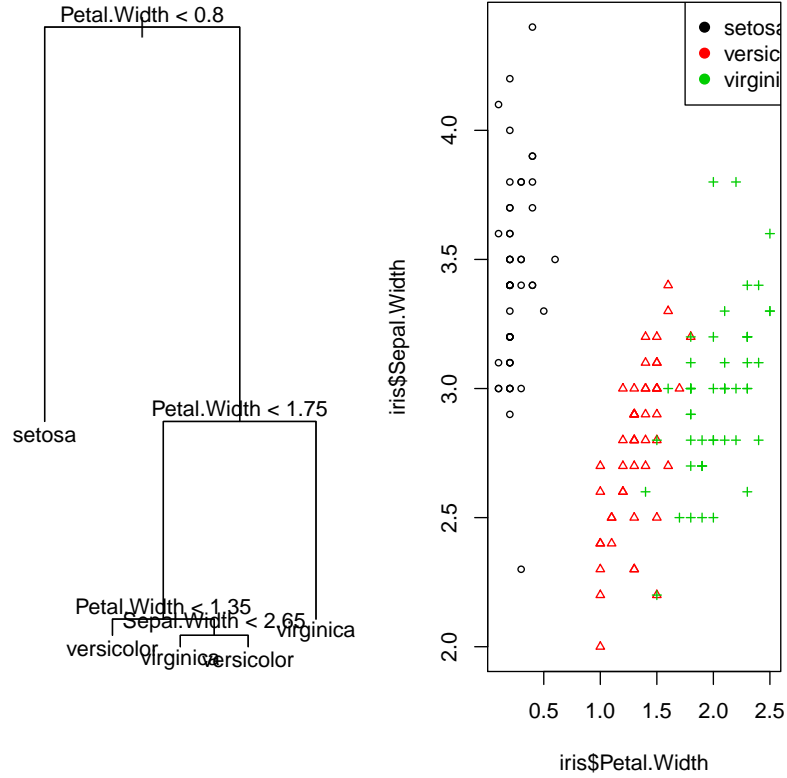
Figure 4: Decision tree model on 'iris' data

rather than from experts. It potentially transforms the biopolitical rendering of differences through specific attributes of individual and populations into a rather mobile matrix of potential mixtures and overlaps.

Take the case of the `iris` dataset, one of the most famous datasets in the machine learning scientific literature. The eugenicist statistician Ronald A. Fisher first used this dataset in his work on the important linear discriminant analysis technique in the late 1930s (Fisher 1938). The `iris` in some ways innocuously enough epitomises the modelling of species differences via measurements of biological properties (in this, measurements of such things as petal widths and lengths of irises growing in the Gaspé Peninsula in Canada).

The plot on the left in Figure 4 shows the decision tree and the plot on the right shows the three *iris* species *virginica*, *setosa* and *versicolor* plotted by petal and sepal widths. As the plot on the right shows, most of the measurements are well

clustered when plotted. Only the *setosa* petal lengths and widths seem to vary widely. All the other measurements are tightly bunched. This means that the decision tree shown on the left has little trouble classifying the irises. Decision trees are read from the top down, left to right. The top level of this tree can be read, for instance, as saying, if the length of petal is less 2.45, then the iris is *setosa*. Recent accounts of decision trees emphasise this legibility: 'a key advantage of the recursive binary tree is its interpretability. The feature space partition is fully described a by single tree. ... This representation is popular among medical scientists, perhaps because it mimics the way a doctor thinks. The tree stratifies the population into strata of high and low outcome, on the basis of patient characteristics' (Hastie, Tibshirani, and Friedman 2009, 306-7). Decision trees do indeed have a rich medical, as well commercial and industrial history of use. Decision trees and their later variations (such as`C4.5`, the 'top' data-mining algorithm according to a survey of data miners in 2009 (Wu et al. 2008)) are often presented as easy to use because they are 'not unlike the serious of troubleshooting questions you might find in your car's manual to help determine what could be wrong with the vehicle' (Wu et al. 2008,2). While that scenario is unlikely today, especially as Google sends autonomous-driving cars out onto the roads of California undoubtedly controlled by a variety of classifiers such as decision trees, neural networks and support vector machines, the recursive partitioning technique still has a great deal of traction in machine learning practice precisely because of its simplicity.

In `iris`, species differences have an ontological or biopolitical weight. The species differences that the decision tree algorithm finds also exists in the world for us. But the recursive partitioning algorithm that constructs the decision tree knows nothing of these differences. Moreover, in many cases, differences are not easily identifiable even for experts. Often there is some pattern of separation, perhaps in the form of overlapping clusters or clouds of points, but not enough to define a simple set of decision rules. How then does a decision tree decide how to split things? Choosing where to cut: this is a key problem for classification or decision trees. What counts as a good split has been a long standing topic of debate in the decision tree literature. As the statisticians Malley, Malley and Pajevic observe, 'the challenge is to define *good* when its clear that no obviously excellent split is easily available' (Malley, Malley, and Pajevic 2011, 121).

Regardless of how 'excellent splits' are defined, the recursive partitioning decision tree subtly alters the nature of decision. Just as Monte Carlo simulation renders all events as probability distributions, or the Perceptron configures classification as on ongoing process of optimisation, decision trees and the like render expert judgment as an algorithmic problem of finding good splits or partitions in data to account for differences. Decision trees, especially in their more recent incarnations of ensembles of decision trees (for example, random forests), no longer formally express the rules that Cold War rationality used to stabilise power-knowledge. The generation of the rules that define the branches in diagrams such as Figure 4 no longer relies on domain experts. Decisions diverge somewhat from cognitive skill or technical rationality to reappear as recursive

14

generated partitions. To give a brief indication of how the decision tree has dispersed in the world, we might think of the Microsoft Kinect game controller, a popular gaming interface that uses video cameras and decision tree algorithm to learn to classify players' gestures and movements, and thereby allow them to play a computer game without touching buttons, levels or holding a game controller. The decision algorithm, coupled with the imaging system, translates the mobility and fluidity of gesture into a set of highly coded movements in the on-screen space of the game. There is little human gesture expertise implemented in the Kinect game controller, only algorithms that learn to classify gestures into a number of categories by propagating gestural date down a decision tree. These categories still refer to conventions and codings of the world, but these classifications have little of the ontological or biopolitical depth, since they only derive from relative density and sparsity in the data.

Conclusion: the pluri-dimensional space

We are dealing here with classificatory operations that differ quite radically from either the artificial intelligence of the Cold War with its attempts to develop computer intelligence, or the expert decision support systems of the 1970s and 1980s that sought to capture domain expertise in software (Collins 1990). Both still assumed that there could be in principle separation or discrimination underpinning decisions and classifications and that algorithmic operations should learn to mimic the discriminative finesse of experts. Cold War rationality remained attached to a decisionistic logic that classification and regression trees, even if only by their sheer abundance, radically re-scales and diversifies.

Monte Carlo simulations and the subsequent Markov Chain Monte Carlo simulations, the Perceptron and its ramification in contemporary neural nets, or the decision tree and its proliferation in random forests and similar ensembles are just some of the matrices of transformation playing out in an operational field spanning contemporary science, media and government. They exemplify diagrammatic operations flowing through a vast and complex space of scientific knowledges, techniques, technologies, infrastructures, and disciplines concerned with pattern and the production of pattern. They are hundreds of other techniques in these spaces, and literally thousands of implementations and variations: Gaussian mixture models, gradient boosted trees, penalized logistic regression, AdaBoost, expectation maximization, linear discriminant analysis, topic models, principal component analysis, independent component analysis, etc. (see (Hastie, Tibshirani, and Friedman 2009) for a reasonably comprehensive textbook listing).

While the exemplary diagrams I have discussed do not exhaust the spectrum of movements in the operational field, they do point to the some of the principal axes along which many contemporary power-knowledges move as they search for hidden patterns and value in data. What happens to pattern in this power-knowledge nexus? A. N. Whitehead proposed that quantity presupposes

15

pattern:

> Thus beyond all questions of quantity, there lie questions of pattern, which are essential for the understanding of nature. Apart from a presupposed pattern, quantity determines nothing. Indeed quantity itself is nothing other than analogy of functions within analogous patterns (Whitehead 1956, 195).

I have been suggesting that we are experiencing a re-patterning of pattern today at the intersection of probabilistic, optimising and recursive partitioning processes. Each of the diagrammatic operations described above comprehends a proliferation of data and various enumerations that can be quantified and via quantification, subjected to computation. Text, voice, image, gesture, measurement, transaction, and many forms of record and recording have and are being ingested by digital systems as digitised quantities. But the quantities or numbers involved presuppose pattern. The promise of pattern recognition, machine learning or data mining has been predicated on finding patterns in data rendered as number, but the production of data as number, and as massive accumulation of numbers might already derive from the shifts in seeing, differentiating and classifying that pattern recognition, data mining and machine learning introduce to the presupposed patterns. If algorithmic operations do locate patterns in data, this location already presupposes certain kinds of pattern. The differences between Monte Carlo simulation, the Perceptron and the decision tree starkly delineate presupposed patterns that guide the relations between quantity that algorithms uncover, optimise or converge towards. I have framed these differences as diagrammatic operations to highlight their dependence on and inherence to criss-crossing visual, semiotic, mathematical, technical, and infrastructural planes. All of these operations have somewhat diagonal tendencies, which project them across disciplinary boundaries (from physics to gaming media, from psychology to weapons development, from mathematical theory to handheld devices) with sometimes remarkable transcontextual momentum. The diagonal tendencies of, for instance, the decision tree with its indifference to the qualifications of quantity – it traverses different kinds of data very easily – differ from those of the Monte Carlo simulation with its intensive sampling of data spaces generated by accumulated random numbers.

Many different processes and decisions depend increasingly on such diagrammatic operations. If pattern itself takes on a new diagrammatic force, if its asignifying assimilation of differences re-iterates what Whitehead terms an 'analogy of functions within analogous patterns,' then the Cold War problems of simulation, optimisation and classification find themselves concatenated in a new configuration. The few hundred neutrons of Metropolis and his co-authors expand to included hundreds of millions of observations of players on XBox-Live; the few hundred scientific articles classified by Maron's early decision tree (Morgan and Sonquist 1963) expand to include several billion DNA base pairs of a cancer genome whose associations are analysed by random forests at a Google

I/O conference demonstration (Mackenzie 2015b); the simple logical functions that almost choked the development of Perceptrons in the 1960s are inundated by the billions of features that deep learning nets at Youtube and Yahoo pipeline into unsupervised object recognition tasks in online video.

In this ramifying diagrammatic re-distribution of pattern, we can expect transformations of and reassignments of subject positions as once quite localised force-relations become strategies generalized to business, government and science. What counts as individual, what counts as population, what categories or differences matter, and what the terms of decisions are, potentially shift or are re-classified in this generalization of the diagrammatic operations. Since their inception in problems of nuclear weapons design, logistics or cybernetics, techniques flow out of the closed-world spaces of the Cold War labs and research facilities. They become banal devices rather than instruments of a decisionistic elite. In this movement, another space takes shape, a space whose dimensions are practically treated as open-ended, and whose potential expansion animates the massive build out of infrastructures and the intense efforts to scale up and scale down circuitry and electronic devices. We might need to think about how it might be possible to inhabit this space of patterns as these patterns become power matrices of transformation. The three general cases discussed above all suggest ongoing instability in what counts as pattern, and how pattern derives from movements through data.

References

Arendt, Hannah. 1998. *The Human Condition*. Chicago ; London: University of Chicago Press.

Bellman, Richard. 1961. *Adaptive control processes: a guided tour*. Vol. 4. Princeton university press Princeton. http://www.getcited.org/pub/101191710.

Bowker, G. 1993. "How to Be Universal: Some Cybernetic Strategies, 1943-70." *Social Studies of Science* 23 (1): 107–27. doi:10.1177/030631293023001004.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Breiman, Leo, Jerome Friedman, Richard Olshen, Charles Stone, D. Steinberg, and P. Colla. 1984. "CART: Classification and regression trees." *Wadsworth: Belmont, CA* 156.

Collins, Harry M. 1990. "Artificial Experts: Social Knowledge and Intelligent Machines." http://www.citeulike.org/group/884/article/633745.

Cortes, C., and V. Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20 (3): 273–97. doi:10.1023/A:1022627411411.

Dahl, George. 2013. "Deep Learning How I Did It: Merck 1st place interview. no free hunch." http://blog.kaggle.com/2012/11/01/deep-learning-how-i-did-it-merck-1st-place-interview/

Edwards, Paul N. 1996. *The Closed World : Computers and the Politics of Discourse in Cold War.* Inside Technology. Cambridge, Mass. ; London: MIT Press.

Erickson, Paul, Judy L. Klein, Lorraine Daston, Rebecca Lemov, Thomas Sturm, and Michael D. Gordin. 2013. *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality.* University of Chicago Press. https://books.google.co.uk/books?hl=en/&lr=/&id=-b0BAgAAQBAJ/&oi=fnd/&pg=PR5/&dq=erickson+cold+war/&ots=BGlorQcl-D/&sig=Hx5igFgX9mILEE8gR54prrayKaU.

Fisher, R.A. 1938. "The statistical utilization of multiple measurements." *Annals of Human Genetics* 8 (4): 376–86.

Foucault, Michel. 1972. "The Archaeology of Knowledge and the Discourse on Language (Trans: A. Sheridan)." *Pantheon, New York.*

Fuller, Matthew, and Andrew Goffey. 2012. *Evil Media.* Cambridge, Mass: MIT Press.

Hacking, Ian. 1975. *The emergence of probability.* Cambridge ; New York: Cambridge University Press.

Halpern, Orit. 2015. *Beautiful Data.* Durham, N.C: Duke University Press.

Hastie, Trevor, Robert Tibshirani, and Jerome H. Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction.* New York: Springer.

Hayles, N. Katherine. 1999. *How We Became Posthuman : Virtual Bodies in Cybernetics, Literature, and.* Chicago, Ill. ; London: University of Chicago Press.

Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. 2006. "Reducing the dimensionality of data with neural networks." *Science* 313 (5786): 504–7. http://www.sciencemag.org/content/313/5786/504.short.

Mackenzie, Adrian. 2015a. "The Production of Prediction: What Does Machine Learning Want?" *European Journal of Cultural Studies* 18 (2): 429–4445. doi:10.1177/1367549415577384.

———. 2015b. "Machine Learning and Genomic Dimensionality: from Features to Landscapes." In *Postgenomics: Perspectives on Biology After the Genome*, edited by Hallam Stevens and Sarah Richardson, 73–102. Durham, N.C.: Duke University Press.

Malley, James D., Karen G. Malley, and Sinisa Pajevic. 2011. *Statistical Learning for Biomedical Data.* 1st ed. Cambridge University Press.

Markoff, John. 2012. "In a Big Network of Computers, Evidence of Machine Learning." *The New York Times*, June. http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html.

18

Mcgrayne, Sharon Bertsch. 2011. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy.* New Haven, CT: Yale University Press.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. "Equation of state calculations by fast computing machines." *The journal of chemical physics* 21: 1087. http://link.aip.org/link//%3FJCPSA6/21/1087/1.

Metropolis, Nicholas, and Stanislaw Ulam. 1949. "The Monte Carlo method." *Journal of the American Statistical Association*, no. 44: 335−341.

Minsky, Marvin, and Seymour Papert. 1969. "Perceptron: an introduction to computational geometry." *The MIT Press, Cambridge, expanded edition* 19: 88.

Morgan, James N., and John A. Sonquist. 1963. "Problems in the analysis of survey data, and a proposal." *Journal of the American Statistical Association* 58 (302): 415–34. http://amstat.tandfonline.com/doi/full/10.1080/01621459.1963.10500855.

Pickering, Andrew. 2009. *The Cybernetic Brain.* Chicago ; London: University Of Chicago Press. http://www.press.uchicago.edu/ucp/books/book/chicago/C/bo8169881.html.

Rosenblatt, F. 1958. "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological Review* 65 (6): 386–408. doi:10.1037/h0042519.

Solovey, Mark, and Hamilton Cravens. 2012. *Cold War Social Science: Knowledge Production, Liberal Democracy, and Human Nature.* Palgrave Macmillan.

Stengers, Isabelle. 2011. *Cosmopolitics II.* Translated by Robert Bononno. University of Minnesota Press.

Stigler, Stephen M. 1986. *The history of statistics: the measurement of uncertainty before 1900.* Cambridge, Mass.: Harvard University Press.

Vapnik, Vladimir. 1999. *The Nature of Statistical Learning Theory.* 2nd ed. 2000. Springer.

Von Neumann, John, and Oskar Morgenstern. 2007. *Theory of Games and Economic Behavior.* Princeton ; Oxford: Princeton University Press.

Whitehead, Alfred North. 1956. *Modes of Thought; Six Lectures Delivered in Wellesley College, Massachusetts, and Two Lectures in the University of Chicago.* New York, Cambridge University Press.

Wu, X., V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J McLachlan, et al. 2008. "Top 10 algorithms in data mining." *Knowledge and Information Systems* 14 (1): 1–37.