Future Psychohistory

Computation and Humanity

Introduction

In his mid-20th century science fiction trilogy **Foundation**[1], Isaac Asimov used the fanciful science of 'psychohistory' as a literary trellis to weave a story around. The concept was simple yet intriguing: humanity's course could be analyzed, predicted, and even guided using this new science that combined elements of mathematics, psychology, and history. New science was a familiar theme of the time. This was the dawn of the atomic age, the blossoming of the computer age, a time of recent and breath-taking expansion of the cosmological model, and the eve of the genetics revolution. Our world view had been radically re-shaped by Darwin, Einstein, Hubble, Turing, and many others.

Computer hardware has evolved considerably since that time, especially in miniaturization[2] and parallel processing. Hardware advances have led to innovation in software and networking, and multiple information technology feedback loops have manifested themselves (e.g. computer-guided design).

There is not as yet (and may never be) a set of formulae that can chart humanity's course as in **Foundation**. However, it may soon be possible to accurately model a form of psychohistory computationally. At least four paradigm shifts are currently underway, making this possible. They include:

1) almost-free transistors
2) the supplanting of the formula by the algorithm
3) abstracted computing (the 'Cloud')
4) the growing realization that brain=mind

This article is neither an advocacy for any course of action nor a general espousal of Asimov's ideals. Rather, it is only a brief look at the possibility of computational psychohistory.

I must, however, confess a deep admiration for Asimov's life and a love since boyhood of his books.

[1] Consisting of *Foundation* (1951), *Foundation and Empire* (1952), and *Second Foundation* (1953)
[2] Unrelated, but miniaturization was the subject of Asimov's novelization of the 1966 film, *Fantastic Voyage*

The Foundation Vision

August 1941. Hitler's armies were on the march all across Europe. A malevolent, dictatorial, and technologically advanced empire, unlike anything the world had seen since Rome, had quickly emerged. Meanwhile, an ocean away, a quiet America was still months away from Pearl Harbor.

In a small office in New York, a young chemistry grad student, who moonlighted as a science fiction writer, met with his publisher. The main topic of discussion was a possible new project, one that would span at least several stories and maybe even several books. The eventual result of that discussion, first published piecemeal, and always guided by Asimov's own philosophy, was **Foundation**.

Here is a very brief synopsis [spoiler alert].

In the far future, the great Galactic Empire has united virtually every corner of the galaxy into a single political and economic union. At the centre of this empire sits its pinnacle, the planet-city of Trantor. It is, for the most part, a benevolent and peaceful system, where order and tradition have dominion. However, it is in the early stages of decline, and is replete with the trappings: a corrupt, stultifying bureaucracy, pompous aristocracy, and even a few vicious rulers.

A small group of academics on Trantor, the 'psychohistorians', are the only ones who know of this impending collapse and they keep it very secret to avoid sedition charges. They are led by the brilliant mathematician and inventor of psychohistory, Hari Seldon. His goal is to let psychohistory guide humanity through a much shorter, less painful collapse and recovery of only a thousand years, instead of the thirty thousand years of ignorance, brutality, and misery that psychohistory predicts if things are left to play out on their own. This is to be accomplished through the establishment of a 'Foundation' to develop science and rationality to the point where it can eventually replace the dead empire and save the galaxy. The book *Lost Horizon* (Hilton, 1933) tells a similar story with the Tibetan lamasery of Shangri-La playing the part of the Foundation.

The tale then recounts the rise of the noble Foundation, from its humble beginnings on the periphery of the galaxy, to its eventual domination of a good sized section of it. This period is marked by a series of 'Seldon Crises', which correspond to critical junctures in the psychohistory 'Plan'. A small screening room, featuring a projection of the long-dead Seldon, re-activates at such crises, usually to state the obvious or be otherwise innocuous. It is an important precept of psychohistory (and thus the Plan) that the general population is not informed by, consciously guided by, or even aware of the Plan.

One of the strongest themes in **Foundation** is that of economics trumping militarism. Over and over in the tale, a brief flash of silly, goose-stepping empire building is washed away by the relentless forces of trade, work, and brave industry.

One example is the interstellar trader Hober Mallow. He leads a trade effort to one of the hostile worlds that surround the Foundation. Through deft trading and negotiating, he manages to make this potential enemy reliant upon Foundation technological goods (and ongoing maintenance of those trinkets). He is basically a swashbuckling merchant. His exploits get the Foundation past a Seldon Crisis, move it beyond being a mere purveyor of a pseudo-religion, and make him the first of the 'Merchant Princes'.

Another example is what happens on the once magnificent Trantor after the Galactic Empire falls.

After it is sacked, it meets with a rather inglorious fate:

> The survivors tore up the metal plating and sold it to other planets for seed and cattle. The soil was uncovered once more and the planet returned to its beginnings. In the spreading areas of primitive agriculture, it forgot its intricate and colossal past
>
> (Asimov, Second Foundation, ch. 18 para. 4)

No trumpets or flags, no marching armies or imperial grandeur - just mooing cows. This is a parable often told by those who either barely escaped (like a young Asimov and his family) or endured great suffering under Stalin or Hitler. It is the utopian 'little chicken farm' referred to in the opening of Frank Capra's 1937 film version of Hilton's *Lost Horizon*.

Science and rationality are two more of the touchstones of the Foundation. They are forced by circumstance to use baser methods at times to deal with stubborn ignorance. However, the leaders of the Foundation always keep their eyes fixed on the great future when civilization, knowledge, and peace will reign supreme (or they suffer greatly for their navel-gazing).

Asimov was intellectual and reflective, but not at all preachy. His books are usually humorous; sometimes even the non-fiction ones, with protagonists who are charmingly less than perfect.

This is a story of humanity as a great, unstoppable juggernaut, but it is definitely not an analytical, objective tome like 'Das Kapital'. Rather, it is chock-full of colourful heroes and villains, together with a lot of suspense and tricky turns. It is pulp fiction with a message.

Eventually, an unforeseen mutant mental giant called 'The Mule' revives and appropriates the old empire, derails the Plan and conquers the Foundation. This is an important literary tip of the hat by Asimov to statistical improbability.

But soon, the here-to-fore quiescent Second Foundation comes to the rescue. This is a tiny, purely academic group, spawned by the original psychohistorians and situated by Seldon at the 'other end of the galaxy' from the first Foundation. During the centuries that the first Foundation was expanding by means of science, technology, and trade, the Second Foundation was covertly developing mental powers and honing psychohistory into a formidable, almost exact, science. They defeat the Mule through mind-alteration, pick up the pieces of the first Foundation, and restore the Plan. Most impressive, they accomplish all of this while keeping their existence concealed.

A sweet and humble farmer turns out to be at the head of the Second Foundation - a final chuckle from Asimov, likely aimed at all the grandiose dictators of the 20th century.

The Transistor

Mastery of any quantity requires flow control, which is normally accomplished by a gate. This could be a fence gate for livestock, a sluice for water, a bank for money, etc. Gating the flow of electricity was one of the major technological advances leading to the modern world.

The early phase of the study of electromagnetism took place in the 1820s with the work of Ampère and Faraday (Steinle). They had to manually connect and disconnect components to control electricity in their elementary circuits. In 1843, Charles Wheatstone invented the rheostat, a manually controlled mechanical device for adjusting the flow of electricity ("rheo" being Greek for "to flow") (Lytle).

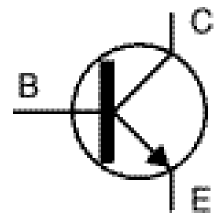
Sir John Ambrose Fleming invented the thermionic electron valve, more commonly known as the vacuum tube, in 1904 (Okamura, 1994). This device allowed the electronic control of current flowing across a vacuum gap. For decades, it served as the basis of amplifying circuits in devices such as radios. It also served as a switching device in early electronic computers as did the electro-mechanical relay.

ENIAC[3], the first full-scale, general-purpose electronic computer, was completed in 1946. It used relays for input/output, and 18,000 vacuum tubes for computation (Norton, 2005, p. 39).

Vacuum tubes were however, large, power-hungry, slow, unreliable, and expensive. The very idea of electrons leaping across a vacuum gap was ungainly. A solid state solution was needed.

In 1947, working at Bell Labs in New Jersey, J. Bardeen and W. Brattain constructed a working transistor (Brinkman, Haggan, & Troutman, 1997).

A transistor is an electronic valve (gate) that controls the flow of current through a solid-state semiconductor (Coles, 1977, p. 5). "A transistor is just a piece of silicon whose conductivity can be turned off and on." (Biermann, 1990, p. 204) Its name is a contraction of TRANSfer resISTOR.



The world now had a small, low-power, fast, reliable, and inexpensive alternative to the vacuum tube. But the best was yet to come – miniaturization.

For several years, designers experimented with different semiconductor types. A breakthrough came with transistors being built using silicon layers (planar construction). Early planar transistors sold for several dollars each around 1959 (Moore, Keynote Speech, 1997).

Using a photonic process, these planar transistors could be produced *en masse* on silicon wafers. Many transistors could be placed together on an ‘integrated circuit’. Dramatic miniaturization took the electronics world by storm, and a dizzying reduction in transistor size and cost began.

[3] Electronic Numerical Integrator and Computer built at the University of Pennsylvania

Moore's law, devised (and later modified) by microelectronics pioneer Gordon E. Moore (Cramming more components onto integrated circuits, 1965), states that the number of transistors that can be squeezed at low cost onto an integrated circuit will double every two years. It is a prediction that has held true to the present[4]. At times, it was even a bit too conservative.

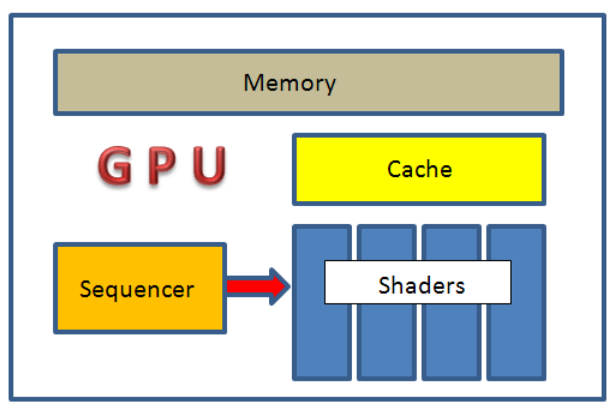
In 1968, the price of a transistor was one dollar (Intel, 2005). By 1972, it was about twelve cents. By 1980, it was a few hundredths of a cent. By 1998, it was one millionth of a cent. It continued into small fractions of a millionth of a cent. Six years ago, it was estimated that the price of a transistor was that of one newspaper character (Intel, 2005). Today, transistors on massively integrated circuits are virtually free. The cost is in the packaging.

Reduction in size/cost is not the only benefit. As integrated circuits get smaller, they also get faster. Electric charge moves through a wire at great speed, but not unlimited speed. The shorter the interconnections between transistors become, the faster will be the operations they perform. For a given speed of electric charge, and a given number of operations per second, there will be a limiting distance that an electric charge can travel. A rough estimate for the speed of electric charge inside an integrated circuit is 1/3000 the speed of light, or 100 kilometers per second (Biermann, 1990, p. 221). Using this number, the limiting distance at 1 thousand operations per second is 328 feet, at 1 million operations per second is 3.9 inches, and at 1 billion operations per second is 0.004 inches (p. 222). Size does indeed matter.

It is difficult to overestimate the impact of this size/cost reduction over the last 50 years. Some of the fruits are: computers (super, mini, micro, portable, video game consoles[5]), communications (satellites, cellphones, microwave, digital TV, networks (including the Internet)), medical (diagnostics, treatment, chemical analysis and synthesis, genetic research), digital cameras, terrestrial navigation (GPS), extraterrestrial navigation (moon race, planetary exploration), manufacturing (assembly/robotics, automotive, aerospace), and new economic growth (Silicon Valley & other high-tech zones, the 'Tiger' economies). Truly, we now live in a digital age.

The GPU

In the mid-1990s, touched off by the burgeoning video game market, secondary pyrotechnics of transistor density began. The graphics processing unit (GPU) was a new dedicated subsystem on a PC card designed to enhance 3D graphic game display. Early cards had about 1 million transistors. By 2000, they had 25 million. By 2005, they had hundreds of millions (Lilly, 2009).



[4] Due to relentless innovation and the huge gulf between our scale and the quantum scale, a physical limit which we are now approaching

[5] "the latest Sony PlayStation would easily outpace the fastest supercomputer from the nineties" (Chazelle, 2006, p. 1 para. 5)

Each core in a GPU is a central processing unit (CPU) in its own right. These cores are often called ‘shaders’ to reflect their main task in graphics manipulation, the GPU’s ostensive purpose. Each is capable of running its own program, or thread. These cores can also access shared memory, and are under the control of orchestrating logic that implements a unified parallel processor.

Today, good GPUs have billions of transistors implementing and controlling hundreds of shaders (cores). These cards are often designed to work in tandem with several together in the same PC chassis. So, a high-end gaming PC might have 10 billion transistors and a thousand cores in a single, 1 kilowatt box. Compare this with the **Foundation**-contemporary, apartment-sized, 150 kilowatt ENIAC and its 18,000 vacuum tubes.

GPU software also evolved. Early proprietary application programming interfaces (APIs) gave way to the Open Graphics Library (OpenGL) (Lilly, 2009). Scientific research and engineering applications began to appear to harness some of this vast new number crunching power too. GPU-assisted research projects in astronomy, biology, genetics, physics, materials/design, mathematics, and others have been launched.

For many applications, computational costs are a receding concern. Almost-free transistors are currently moving several pre-existing techniques from the mostly academic realm into everyday practicality. We now look at two of these: Bayesian learning and functional programming.

Bayesian Learning

Bayes[6] introduced a method for combining a prior probability together with new knowledge (results data) to infer a revised probability. In Bayesian inference, subjectivity is made explicit and fixed in the past. Different sources may supply different prior probabilities (possibly subjective beliefs). This is in contrast with other statistical methods which may seem more objective, but where in fact the subjectivity is made implicit and on-going (e.g. assumptions about randomness, choices about sampling).

The revised probability can then be fed back into the algorithm as the prior probability, together with new data. As this process iterates, learning takes place.

Bayes' theorem[7]:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

where:

P(H) is the prior probability of hypothesis (existing belief)
P(D|H) is the conditional probability (likelihood) of data given hypothesis
P(D) is the probability of data
P(H|D) is the posterior probability of hypothesis given data (revised belief)

For two centuries, Bayesian theory was frowned upon by most statisticians, and was the source of a deep schism. The disagreement is essentially rooted in one's definition of probability.

For a 'Frequentist', an event's probability is measured by the frequency distribution of trial outcomes. If an event has a probability of 0.5, it means that it is equally likely as not to occur in the limit (infinite trials). At the heart of this viewpoint is implicit randomness. Subjectivity about causes is excluded from the process.

For a 'Bayesian', probability is a means of quantifying lack of knowledge (degrees of belief). New knowledge is used to revise beliefs. Effects inform knowledge about causes. It is an inductive process of learning from evidence or experience. Any subjectivity is made explicit, but past (prior), where it is more difficult to argue subjectivity.

Far from excluding or avoiding causality, causal Bayesian networks explicitly include causes. "One of the most intriguing aspects of Bayesian networks is the role they play in formalizing causality." (Darwiche, 2010, p. 88)

The use of prior distributions in Bayesian methods is a double-edged sword. To a Frequentist, it is an unnecessary and subjective complication. To a Bayesian, however, it is an opportunity to include more information into an inference.

Perhaps an exclusive distinction between the two is unnecessary. "The man in the street is happy to use probabilities in both these ways." (MacKay, 2003, p. 26) Laplace said of the latter: "essentially, the theory of probability is nothing but good common sense reduced to mathematics" (as quoted in (McGrayne, 2011, p. 50)) "Of course, it is futile to argue over which is the 'correct' definition of probability. The different definitions merely reflect different choices for the types of problems the theory can address" (Loredo, 1990, p. 85)

A problem in applied mathematics is the 'curse of dimensionality'. This refers to the idea that some problems ('some' becomes 'most' when studying examples from nature) have too many variables to

[6] Thomas Bayes 1702-1761. English Presbyterian minister, theologian, and mathematician (EB, Thomas Bayes, 2011)
[7] The Modern representation is due to Laplace (McGrayne, 2011, pp. 49-50)

be consistently handled. Worse, these variables often influence each other. Explicitly handling each and every eventuality could lead to a combinatorial explosion which would easily overwhelm the most massive computational power available now or even in the future.

The brain is one such example from nature. The frequentist approach, which applies best to problems with a few un-entangled variables, is hopelessly out-gunned. The Bayesian approach may still have a chance:

> for multivariate cortical data, the Bayesian model provides for a more accurate representation by removing the effect of confounding correlations that get introduced due to canonical dependence between the data
>
> (Joshi, Joshi, Leahy, Shattuck, Dinov, & Toga, 2010, abstract)

Another problem from nature is that of mapping RNA onto DNA. Translation from DNA to RNA involves keeping certain sequences (exons) and removing others (introns). This makes mapping a 'finished' gene back onto DNA problematic. One approach is to use canonical Bayesian networks to study genetic linkage:

> To assess the likelihood of a linkage hypothesis, one uses a pedigree with some information about the genotype and phenotype of associated individuals. Such information can be systematically translated into a Bayesian network
>
> (Darwiche, 2010, p. 85)

Sometimes, we need to evaluate probabilities with rare phenomena. Frequencies do not apply:

> The gamma ray astronomer does not want to know how an observation of a gamma-ray burst would compare with thousands of other observations of that burst; the burst is a unique event which can be observed only once, and the astronomer wants to know what confidence should be placed in conclusions drawn from the one data set that actually exists
>
> (Loredo, 1990, p. 83)

Any model of real-world complexity is far beyond any engineer's or artist's ability to manually design. Such complexity must be created dynamically and/or *learned* from real-world data. Bayesian methods are an excellent means of implementing machine learning and subsequent machine reasoning and understanding that can generate new complexity based on the model.

Why did Bayesian analysis languish for centuries? Dogmatism and the success of the Frequentist approach (at least until recently) are indeed part of the reason, but there is also a simpler part. Paradoxically, the world had to catch up to Bayes and Laplace. The computational power required for Bayesian methods to be successfully applied to real world problems has only recently existed:

> Computations took forever…
> The title of a meeting held in 1982, "Practical Bayesian Statistics", was a laughable oxymoron.
>
> (McGrayne, 2011, p. 245)

Functional Programming

In 1941, as Asimov ruminated over **Foundation**, Alonzo Church wrote "The Calculi of Lambda-conversion". His lambda calculus provided a conceptual basis for the discipline of functional programming (Barendregt, 1997). FP, as opposed to the more common imperative programming approach (e.g. C), is well-suited to concurrent and parallel architectures. There are now many FP languages and tools, and FP has also been included at least to some degree in many other popular languages and tools.

We will take a closer look at one functional programming language, Haskell, in order to highlight some of the features of FP that make this a promising approach for modern scientific modeling. Haskell is certainly not the only, or necessarily the best, choice[8]. However, it is by default: purely functional, 'lazy', compiled, and statically (although automatically) typed. It is open-source, has been in active development for over twenty years, and has built-in support for parallel systems (Haskell.org).

In 1958, John McCarthy, inspired by the lambda calculus, created Lisp, the main language of artificial intelligence (AI) for decades to come. In the early 1970s, Robin Milner created the functional programming language ML. Haskell 1.0 appeared in 1990. Rapid increase in computer power allowed trading off some raw performance in favour of programmer productivity. Haskell then successfully made the move beyond academia and into the open source and commercial domains (O'Sullivan, Stewart, & Goerzen, 2008) (A brief sketch of Haskell's history section).

Haskell is 'purely functional'. This means that Haskell functions are without side effects. Code execution does not need to be sequential because variable data are not passed along a chain as on an assembly line. The classic statement X = X + 1 has no place in Haskell because the data that a function works with are never modified. Access to the external world is not handled by 'pure' code. Automated testing of purely functional code is easier because it is 'stateless' (immutable data) and isolated from the outside world.

Haskell is very high level (i.e. elegant, powerful). Its syntax is less like a series of explicit steps and more like mathematical expressions. One specifies 'what' is needed, not 'how' to get it (declarative as opposed to imperative). Since it is stateless, FP implements repetition via recursion not loops (Goldberg, 1996).

Haskell is 'lazy'. This means that functions are not executed unless and until they are required, improving brevity and efficiency, sometimes immensely. The value of blinding execution speed is lessened when a language is savvy enough to avoid wasted effort.

Haskell is 'rigorous'. By design, errors are detected at compile time, not runtime. Actually, a program's correctness is almost a prerequisite to it running at all.

Since it is purely functional, Haskell code runs well on **parallel** architectures. Since it is compiled, statically typed, and 'lazy', Haskell code is **fast**. Since it is elegant, rigorous, and highly testable, Haskell code is **maintainable**. This combination makes Haskell suitable for modeling huge systems such as the psychohistory of humanity.

[8] For example, Clojure springs to mind, but the requirement for a Java VM *might* be problematic.

The Algorithm

Formal logic has been a shining jewel for great thinkers down through the ages. In the West, one can list Aristotle, Hobbes, Leibniz, Boole, Gödel, and many others before even venturing into the rich philosophical histories of other cultures.

For millennia, mathematics progressed based on symbolic representations and formal logic. As a result, our understanding of the universe also progressed. Perfect shapes and exact formulae inspired mathematicians, astronomers, physicists, engineers, artists, and philosophers to embrace the world of axioms, conjectures, proofs, and truth exemplified by this 'Queen of Sciences'.

In the last century or so, as in so much of science, storm clouds have begun to appear. We have been increasingly choked by complexity and data. Now faced with multivariate, nonlinear, near-frantic complexity, fed to us by instruments that extend our reach far out in all directions of time, space, and life, our beloved symbolic notions have begun to fall short.

As it turns out, there is another way; one that is unfettered by the requirement of fitting within the human head.

> The Algorithm's coming-of-age as the new language of science promises to be the most disruptive scientific development since quantum mechanics
>
> (Chazelle, 2006)

> The younger generation seemed to think that computers and their algorithms could replace mathematics entirely.
>
> (McGrayne, 2011, p. 245)

Alan Turing wrote a paper in 1936 which answered a question in mathematics (negatively) by replacing calculus with a 'universal computer' (Steer, Birch, & Impney, 2008, p. 259). Here was a theoretical outline of the modern computer.

The universal computer was thought objectified. Any possible task of computation could be done by this theoretical contrivance. Over the next decade, theory would become practice, and the ENIAC appeared in 1946.

Computers got faster, smaller, and cheaper, but the basic principles of operation have remained unchanged since Turing[9]. A central processing unit (CPU) reads programs and data[10], processes them ('control'), and produces output.

A formula is a predicate applied to one or more arguments, or a combination of simpler formulae (Sharples, Hutchinson, Torrance, & Young, 1989, p. 361). "An algorithm is a method, procedure, or recipe for doing a job." (Biermann, 1990, p. 39)

The algorithm's procedure is applied by the controlling logic (CPU) as it reads input and (presumably) generates output. A formula relates symbols, an algorithm manipulates data. Overly simplified, a formula *articulates* knowledge, while an algorithm *generates* knowledge (or at least information).

Algorithms are now in use everywhere, perhaps even in your digital camera. Some of the earliest and most successful algorithms were invented for image processing. Edge and region detection were implemented years ago (Sharples, Hutchinson, Torrance, & Young, 1989, p. 266).

---

[9] Excepting theoretical ideas like quantum computing
[10] The similarity between programs and data is known as 'duality' (Chazelle, 2006)

One of the mathematical forms that has come to the fore because of ever increasing computational power is the fractal. The best-known fractal is the Mandelbrot set, which is generated with the algorithm:

$$Z \to Z^2 + c$$

where:
Z is a complex number
c is a complex constant (test point)

After sufficient iterations, a planar image emerges displaying self-similarity at different scales and an "infinite regress of detail" (Dewdney, 2002). The Mandelbrot set illustrates that a simple algorithm using basic iteration or recursion can produce deep complexity. Fractals offer a glimpse into the construction mechanisms of nature.

Another form is the cellular automaton, which is again the basis of many simulations of systems in the real world. The best-known example is Conway's Game of Life (EB, cellular automata (CA), 2011).

Another powerful method, especially useful for research in areas where little is known about causation, is Monte Carlo analysis. In this method, random sampling is used instead of a formula. The goal is to simulate a complex system and to have properties of that simulation converge to stable values. In normal cases, Monte Carlo analysis helps to test theory. In the best case scenario, actual discovery of underlying causation is a goal (to guide theory).

A very simple example of Monte Carlo analysis is the estimation of the value of π. The following diagram shows the geometric and algebraic representations of the ratio of areas of a circle (radius=r) which fits exactly inside a square (side=2r). Thus, knowing this ratio will give us a direct means of calculating π.

$$\frac{Acircle}{Asquare} = \pi r^2/(2r)^2 = \pi/4$$

We repeatedly generate random points that lie within the square. Some of these points will also lie within the circle. By counting the number of 'hits' in both, we can estimate the ratio of areas of the circle to the square. Generally, the more random points (samples) are generated, the more accurate the estimate will be. It can take a million samples for the value of π to converge to several decimal places. However, this task can be accomplished by a modern parallel processor in a fraction of a second. Computational costs are a receding concern.

A more sophisticated method is Bayesian Monte Carlo (BMC) which incorporates prior knowledge. (Rasmussen & Ghahramani, 2002, abstract)

Algorithms offer informal and highly interactive access to knowledge manipulation and testing. They encourage inductive thought as opposed to, and hopefully in addition to, more traditional, logical deductive reasoning. They have become a powerful tool for scientific thought, research, and even discovery.

"One can't proceed from the informal to the formal by formal means." [11]

[11] Computer scientist Alan J. Perlis

The Cloud

Classical computing involved the local assembly and maintenance of computing resources and expertise. It was the only realistic approach until about a decade ago. The Internet, coalescence of standards (e.g. security, formats, and workflows), and the maturing of specialized services has led to a modern alternative.

Cloud computing is the accessing of services delivered over the Internet. Computing is abstracted to a virtual space called the 'Cloud'. This approach enables small organizations (e.g. many research institutions) or even individuals to obtain access to significant processing power. Additionally, they can benefit from expertise and experience that would be too costly, difficult, and time-consuming for them to develop themselves. This includes development, maintenance, and updating of software and equipment[12].

Cloud computing is infinitely scalable. Resources can be sourced from a global pool. These resources can be quickly made available for special events, projects, or training, and subsequently scaled back again.

As newer hardware, software, and standards arise, they can be capitalized on immediately with little or no buy-in cost. Forecasting the future of technology is no longer a major concern.

This decoupling of research administration from technology is helpful for two reasons.

The first is the simplification and clarification that ensues. When computation is an abstract layer, focus can return to research. Power transmission technology has similarly geographically decoupled electrical power generation from its use.

[12] An example is the Xen virtualization platform developed in the UK. (Schubert, 2010, p. 37)

The second reason is the fact that technology, particularly commercial technology, has always been rife with irrational and stubborn loyalties. Although the same could be said for research at times, science has a self-correction mechanism (e.g. ending the dogmatic dismissal of Bayesian methods) that is lacking in the commercial world where the rationality of purchases is often immaterial.

Distributed Computing

A special case of abstracted computing, where processing is spread widely across geographical boundaries is that of distributed computing (DC). Just as multiple cores in a parallel processor shorten the time it takes to accomplish a task otherwise handled by a single CPU, multiple sites of computation can leverage work by incorporating the processing power of other machines. In most cases, the wider participation in and contributions to projects is volunteered, often by the general public.

Crowd Sourcing

In crowd sourcing, it is not distributed computers that are harnessed, but people, or more specifically, their ideas and opinions. If an open call is put out for participation in a project, it is likely that mainly those with an interest or expertise to offer will reply, thus improving upon a purely random sample.

Multiple distributed human minds can be applied to a common project analogously to distributed computers. Social networks have themselves become a subject of Bayesian learning research. (Acemoglu, Dahleh, Lobel, & Ozdaglar, 2010)

Brain and Mind

Cause: Life

Darwin presented us with a scientific theory of life: evolution[13]. Evolution largely explains the bewildering diversity and complexity of life. As an equation with operators, it might be written as:

```
evolution = time(natural_selection(variation))
```

```
where 'variation' is used as a simplistic
amalgamation of mutation, genetic drift,
geographic isolation, symbiotic combination,…
```

"Natural selection is an improbability pump: a process that generates the statistically improbable." (Dawkins, 2009, p. 416)

Life began almost 4 billion years ago, after the new-born earth had stabilized. "In effect, life on earth began almost as soon as it could have" (Hunter, 1993, p. 8). Evolution went straight to work on simple life forms.

About 1½ billion years ago, cells with a nucleus appeared, and this more complex form of life probably formed from symbiotic collections of simpler, prokaryotic life as described by the Endosymbiotic theory of Lynn Margulis (as cited in Hunter, 1993, p. 8). Then came intra-form cellular specialization (differentiation) (e.g. roots and leaves) (p. 10).

Eventually, immune systems developed, and the power of evolution began operations on the micro-time scale, in a biological analog of the transistor story.

Darwin never knew about DNA. Like digital computers, genetics is based on a digital code. "The machine code of the genes is uncannily computerlike." (Dawkins, 1995, p. 17)

The translation of this code into physical form is called gene expression. The intricate process of gene expression is controlled by "an elaborate dance with many participants" (Hunter, 1993, p. 12). This process is fundamental to the study of molecular biology[14].

Simple life is not 'primitive'. It is evolving. Human life is not 'ultimate'. It is evolving. Complex life evolved from earlier simpler forms, but not from modern simpler forms. All forms of life, from simple to complex, continue to evolve. Evolution is not a single sequential process, but an entire world of sequential processes, sometimes overlapping and even interacting, and all happening in parallel. The temporal mapping of this grand parallel process is a truly immense hierarchical structure - the tree of life.

Long ago, along one branch (or several) of that tree, brains evolved. Eventually, along a subsequent branch, primate brains evolved. Then, along a subsequent branch to that, the modern human brain evolved.

Effect: Brain

Perhaps the most amazing thing about the brain is its size. Confucius, Aristotle, da Vinci, Laplace, and Einstein each perceived the world, and changed it, using this small lump of flesh. 25 years ago, Asimov the biochemist described it conservatively:

> In its three pounds are packed ten billion nerve cells and nearly one hundred billion smaller supporting cells.
>
> (Asimov, Foreword, 1986)

[13] Later work by Fisher, Haldane, and Wright led to a more rigorous description of natural selection, called Neo-Darwinism (Steer, Birch, & Impney, 2008, p. 128)

[14] DNA -> RNA -> protein (Steer, Birch, & Impney, 2008, p. 274)

The brain is like the Endosymbiotic theory writ large. As with all other organs, the brain is composed of cells, but these cells form a vastly interconnected, self-controlled, single entity with complexity and behaviour qualitatively beyond its constituent parts.

Most of the higher human brain functions are located in the Neo Cortex, the crinkled wrapping of the brain, roughly the size (when unfolded) of a dinner napkin, containing about 30 billion neurons (Hawkins, 2010).

The brain looks Bayesian: “In this debate, there is no more powerful argument for Bayes than its recognition of the brain’s inner structures and prior expectations.” (Stuart Geman as quoted in (McGrayne, 2011, p. 286))

In modern theory, brain function is hierarchical. That is, inputs from the senses travel up a pyramid-like hierarchy of neural levels as they self-reference and coalesce into higher and higher ‘thoughts’. The locomotion of this travel is of course the firing of neurons. Near the top of the hierarchy, thoughts enter into consciousness.

Jeff Hawkins describes ‘Hierarchical Temporal Memory’ (Hawkins, 2010) and Michael Shermer describes a ‘binding’ process (Shermer, 2011, pp. 115-117).

It is tempting to assume that ‘hierarchy’ implies design. This is false. Natural hierarchies can arise due to self-organization, evolution, or inductive learning. Hierarchies are not always deduced.

It is also tempting to assume that 100% of intelligence is brain-based. This is unlikely. Neurons do exist outside of the brain, and intelligence might well exist outside of neurons.

Brain's Effect: ‘Mind’

Dualism, the belief that the brain is physical and the mind is not, is as old as mankind. “the French philosopher René Descartes (1596-1650) believed that humans were guided by an immaterial mind” (Sharples, Hutchinson, Torrance, & Young, 1989, p. 9). Descartes[15] lived three centuries before modern neurology so it is understandable that he believed the brain and the mind to be separate.

“They are not. They are one and the same.” (Shermer, 2011, p. 153)

Many people believe that science is limited, and that mind and imagination are ultimately larger. Counter-intuitively, the opposite is true.

We once looked up at the night sky and imagined mythical creatures. We now know that we are seeing photons emitted many years ago by incredibly distant, raging nuclear fusion infernos. Big telescopes reveal soul-drenching beauty and robot probes visit other worlds.

We used to imagine that disease had mysterious dominion over us and that occasional recovery was miraculous. We now have medical knowledge that routinely saves children and extends lives.

Now we are faced with the possibility that the brain, culture, and humanity itself are all valid domains of scientific research, and may even be computable. Predictably, we recoil at first. The ‘masters of the Earth’ now seem like little children wandering into a great library. But perhaps there is grandeur in this view…

The world is not only stranger than we imagine, it is stranger than we *can* imagine.[16]

[15] “I think, therefore I am.”
[16] Derived from original quote by J.B.S Haldane

The Model

The above digression into neurology did have a purpose. We are not considering a normal computer model. This is not a model of particles and force fields, matter and energy, economic interplay, or even complex biochemistry. It is rather, a model of myriad perceptions and beliefs – a model of *models*.

A quick overview of Asimov's psychohistory

Psychohistory began with a rather simplistic view:

> Implicit in all these definitions is the assumption that the human conglomerate being dealt with is sufficiently large for valid statistical treatment. The necessary size of such a conglomerate may be determined by Seldon's First Theorem which ... A further necessary assumption is that the human conglomerate be itself unaware of psychohistoric analysis in order that its reactions be truly random…
>
> (Asimov, Foundation, ch. 4 para. 2)

So Seldon's original view was of people being almost like molecules in a kinetic theory of gases.

However, subsequent books depended on psychohistory taking centuries of epic mathematical effort to substantially develop. As the saga continued, psychohistory became steadily less simplistic and static, and thus more interesting:

> So he [Seldon] created his Foundations according to the laws of psychohistory, but who knew better than he that even those laws were relative. He never created a finished product. Finished products are for decadent minds. His was an evolving mechanism…
>
> (Asimov, Second Foundation, ch. 6 para. 20)

The third sentence of that passage could well serve as a one-line summary of Asimov's philosophy in **Foundation**.

Much later, in the final book of the larger story, Asimov even had one protagonist say:

> the whole thing depends on dealing with people who are both numerous and unaware. Doesn't that seem to you a quicksandish foundation on which to build an enormous mathematical structure?
>
> (Asimov, Foundation and Earth, ch. 31 para. 47)

A quick overview of computation in 2011

Traditionally, computer models have been implemented on sequential computers, using a single central processing unit (CPU). The easiest way of augmenting these models is by extending into parallel hardware architectures like the GPU. Almost-free transistors and cloud computing make this extension relatively easy and inexpensive.

Similarly, parallel software architectures like functional programming make sense. This is now a mature technology, and mature maintenance features enable its use in very large projects. Today's software designers are comfortable with a less formal, more empirical, algorithmic world.

Now that Bayesian methods are more widely accepted and understood, their application should provide the scalability required to conduct quantitative research into nature.

Consciousness is learned

What is the essential trait that makes us human? It certainly is not our senses - any child can quickly think of animals that are more acute in one or more senses. Likewise for our physical abilities. Other creatures have larger brains (e.g. elephants, some whales). Neanderthals may even have had slightly bigger brains, and they are extinct (as a separate species at least).

Yet in a momentary flash, a vanishingly tiny fraction of evolutionary time, modern humans have come to dominate the planet completely, have taken the first few steps into space, and have even begun to manipulate life at the genetic level. How?

This a crucial question if we are to model psychohistory. Any successful model must understand and incorporate the behaviour of its elements. In a psychohistory model, behaviour on every scale depends on the answer to this question, from the individual, to inter-personal relationships, to extended families and circles of friends, to cities, to states, to cultures, to all of humanity.

Of course, the answer is: language.

"Some philosophers believe that it [consciousness] is crucially bound up with language, which seems to have been achieved once only, by the bipedal ape species Homo sapiens." (Dawkins, River Out of Eden, 1995, p. 157)

The study of language is linguistics. Our model must possess considerable expertise in linguistics. In fact, it must be able to acquire/develop such expertise. Since at least the 1950s, the field of computational linguistics has been very active:

> In September of 1952, I presented an oral report…
> As the years have passed, the general awareness that the linguistic problems…
>
> (from an early paper on machine linguistics in a chemistry application) (Garfield, 1961, p. 458)

A model that learns

Psychohistory is not finished and static, it adapts and evolves. Our model must be able to 'learn'.

Early AI focused on symbolic reasoning (math, chess). But that approach hit a brick wall when later research turned to the more everyday problems of vision and common sense. Intelligence is a function not of processing speed, but of knowledge capacity and representation.

As it turns out, one almost has to understand the world before one can understand the world. A person is capable of common sense because they carry around a tiny model of reality in their head.

This model is trained in a 'bootstrapping' manner. When we are infants, we have a very sparse model and can interpret only some simple inputs (a few faces and a few physical mechanisms). Several months spent training that model (mainly at the bottom end of the neural hierarchy) gives us a model capable of taking the next incremental steps. During a lifetime of learning, the model builds on what it already knows to assimilate more information and more complexity. Obviously, there is no objective 'trainer'. A learning mind is a model that self-trains.

A psychohistory model must have this same ability. But instead of one mind, it must train billions[17], while it is simultaneously learning the patterns of causation for humanity's collective behaviour. This model would be 'the mother of all neural networks'.

A few tips & tricks

The only representational architecture that could work for such a huge knowledge base of such enormous complexity is hierarchical. Recursive reuse of stored knowledge and self-reference must be automatic and ubiquitous.

[17] Using knowledge of geography, history, basic human cognition, psychology, language, culture, …

The model should embrace paradox and irony. In cybernetics terms, it should be able to digest a double-bind dilemma.

The model should leverage modern techniques with a healthy respect for the past:

> Current strategies for designing computers that could perform at biological levels exploit such ancient principles as reusable parts, hierarchical structures, variations on themes, and regulatory systems.
>
> (McGrayne, 2011, p. 286)

The designers should learn from others who have taken the first steps into Bayesian modeling of mass social behaviour.

> The social network consists of the network topology and the signal structure. Each individual then chooses one of two possible actions depending on his posterior beliefs given his signal and the realized neighborhood.
>
> (Acemoglu, Dahleh, Lobel, & Ozdaglar, 2010, p. 16)

The overall problem may be somewhat reducible. One avenue might be indicated by 'mirror neurons'. These are neurons that fire when behaviour is *observed in another*. They do not only participate in patterns - they *reflect* them. This might lead to a way to substitute reference for instantiation, an optimization practice familiar to computer programmers.

> there are neurons specialized for discriminating between different intentions … this implicates mirror neurons in both predicting others' actions and inferring their intentions
>
> (Shermer, 2011, p. 132)

Bayesian networks have other major benefits. They can use and reuse many existing generalized algorithms (both for knowledge capture and optimization). They can efficiently represent huge knowledge bases, in part due to their ability to *localize causality* and thus un-entangle variables.

> every variable in the structure is assumed to become independent of its non-descendants once its parents are known
>
> (Darwiche, 2010, p. 82)

Lastly, there are times when traditional smoothing and approximation still can be helpful to simplify Bayesian representations. An example might be to strategically apply a Gaussian process, with one caveat being the computational cost of inverting large matrices (MacKay, 2003, p. 547). However, computational costs are a receding concern.

Conclusion

Over 60 years ago, Isaac Asimov wrote about psychohistory as a calculus for the course of humanity.

In the near future, it may be possible to accurately model a form of psychohistory.

References

Acemoglu, D., Dahleh, M. A., Lobel, I., & Ozdaglar, A. (2010). Bayesian Learning in Social Networks. *Review of Economic Studies, (2010) 01*, 1-34.

Asimov, I. (1951). *Foundation.* New York: Gnome Press.

Asimov, I. (1952). *Foundation and Empire.* New York: Gome Press.

Asimov, I. (1953). *Second Foundation.* New York: Gnome Press.

Asimov, I. (1986). Foreword. In J. Hooper, & D. Teresi, *The 3-Pound Universe.* New York: St. Martin's Press.

Asimov, I. (1986). *Foundation and Earth.* New York: Doubleday.

Barendregt, H. (1997). The impact of the lambda calculus in logic and computer science. *The Bulletin of Symbolic Logic, 3*, 182.

Biermann, A. W. (1990). *Great Ideas in Computer Science.* Cambridge, Mass.: MIT Press.

Brinkman, W., Haggan, D., & Troutman, W. (1997). A history of the invention of the transistor and where it will lead us. *IEEE Journal of Solid-State Circuits, 32*(12), Abstract.

Chazelle, B. (2006). *The Algorithm: Idiom of Modern Science.* Retrieved 05 10, 2011, from http://www.cs.princeton.edu/~chazelle/pubs/algorithm.html

Coles. (1977). *TRANSISTORS: Theory and Use.* Toronto: Coles Publishing Company Limited.

Darwiche, A. (2010). Bayesian Networks. *Communications of the ACM, 53*(12), 80-90.

Dawkins, R. (1995). *River Out of Eden.* New York: BasicBooks.

Dawkins, R. (2009). *The Greatest Show on Earth: The Evidence for Evolution.* New York: Free Press.

Dewdney, A. K. (2002). *The Mandelbrot Set.* Retrieved 06 07, 2011, from University of Waterloo Faculty of Mathematics: http://www.math.uwaterloo.ca/navigation/ideas/articles/mandelbrot/index.shtml

EB. (2011). *cellular automata (CA).* Retrieved from http://www.britannica.com/EBchecked/topic/862593/cellular-automata-CA

EB. (2011). *Thomas Bayes.* Retrieved from http://www.britannica.com/EBchecked/topic/56807/Thomas-Bayes

Garfield, E. (1961). An Algorithm for Translating Chemical Names to Molecular Formulas. *Essays of an Information Scientist, 7 - 1984*, 441-513.

Goldberg, B. (1996). Functional programming languages. *ACM Computing Surveys, 28*(1), 249.

Haskell.org. (n.d.). *The Haskell Programming Language*. Retrieved 06 08, 2011, from haskell.org: http://www.haskell.org

Hawkins, J. (2010, March 18). Hierarchical Temporal Memory [Video file]. UBC, Vancouver, British Columbia, Canada: Retrieved from http://www.youtube.com/watch?v=TDzr0_fbnVk.

Hilton, J. (1933). *Lost Horizon.* London: Macmillan.

Hunter, L. (1993). *Molecular Biology for Computer Scientists.* AAAI Press.

Intel. (2005). *Moore's Law.* Retrieved 06 07, 2011, from Intel Museum: http://download.intel.com/museum/Moores_Law/Printed_Materials/Moores_Law_2pg.pdf

Joshi, A., Joshi, S., Leahy, R., Shattuck, D., Dinov, I., & Toga, A. (2010). Bayesian approach for network modeling of brain structural features. In R. C. Molthen, & J. B. Weaver (Ed.), *Medical Imaging 2010: Biomedical Applications in Molecular, Structural, and Functional Imaging.*

Lilly, P. (2009, 05 19). *From Voodoo to GeForce: The Awesome History of 3D Graphics.* Retrieved 06 06, 2011, from Maximum PC: http://www.maximumpc.com/article/features/graphics_extravaganza_ultimate_gpu_retrospective

Loredo, T. J. (1990). From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics. In P. Fougère (Ed.), *Maximum Entropy and Bayesian Methods* (pp. 81-142). Dordrecht, NE: Kluwer Academic Publishers.

Lytle, B. (n.d.). *The History of the Rheostat.* Retrieved 06 07, 2011, from eHow.com: http://www.ehow.com/about_5376550_history-rheostat.html

MacKay, D. (2003). *Information Theory, Inference, and Learning Algorithms.* Cambridge, UK: Cambridge University Press.

McGrayne, S. B. (2011). *The Theory that Would Not Die.* New Haven & London: Yale University Press.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics, 38*(8).

Moore, G. E. (1997). Keynote Speech. *Intel Developer Forum, Fall 1997.* San Francisco: http://www.intel.com/pressroom/archive/speeches/gem93097.htm.

Norton, J. (2005). *From Gutenberg to the Internet: a sourcebook on the history of information technology Volume 2.* Novato, CA: historyofscience.com.

Okamura, S. (1994). *History of Electron Tubes.* Amsterdam: IOS Press.

O'Sullivan, B., Stewart, D., & Goerzen, J. (2008). *Real World Haskell.* Sebastopol, CA: O'Reilly Media.

Rasmussen, C. E., & Ghahramani, Z. (2002). *Bayesian Monte Carlo.* Retrieved 06 07, 2011, from Advances in Neural Information Processing Systems (NIPS): http://books.nips.cc/papers/files/nips15/AA01.pdf

Schubert, L. (2010). *The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010.* European Commission, Information Society & Media.

Sharples, M., Hutchinson, C., Torrance, D., & Young, D. (1989). *Computers and Thought: A Practical Introduction to Artificial Intelligence.* Cambridge, Mass.: MIT Press.

Shermer, M. (2011). *The Believing Brain.* New York: Times Books.

Steer, M., Birch, H., & Impney, A. (Eds.). (2008). *Defining Moments in Science.* London: Cassell Illustrated.

Steinle, F. (n.d.). *Experiment and concept formation in early electrodynamics: Ampère and Faraday.* Retrieved 06 09, 2011, from IZWT - Universitat Wuppertal: http://www.izwt.uni-wuppertal.de/en/AmpereFaraday