Australian Government
Department of Defence
Defence Science and Technology Group

Unpredictable Outcomes in Unstructured Environments

the future of machine reasoning

Dr Darryn Reid
Principal Scientist
Joint & Operations Analysis Division
Defence Science & Technology Group

The third choice of the military enlightenment

"Though our intellect always longs for clarity and certainty, our nature often finds uncertainty fascinating."

"The very nature of interaction is bound to make it unpredictable."

"The difficulties accumulate and end by producing a kind of friction that is inconceivable unless one has experienced war."

"Circumstances vary so enormously in war, and are so indefinable, that a vast array of factors has to be appreciated."

- Karl von Clausewitz

The mechanistic paradox

People tend to believe

- Knowledge is gained by 'objective' observers gorging on data and extracting the truth out of it
- Every event has a necessary and sufficient set of causes
- Every question has a complete and correct answer, and that it is obtainable, at least in principle
- Decisive action hinges on relative certainty
- Future outcomes can be predicted by governing laws and a sufficiently accurate account of the past
- All failure is inherently pathological
- More data gives more information
- Incremental improvements will eventually accumulate into success

... we live in a clockwork managerial universe.

Methodological hallucinations

People also tend to believe

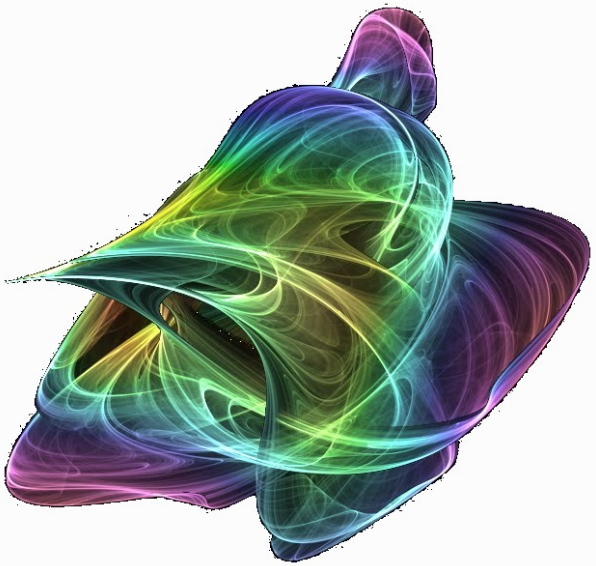
- We can precisely define our problem situations
- We can accurately define complete and correct solutions

- The outcomes are pre-determinable
- The path from problems to solutions is a linear, a matter of efficiency and expected utility
- Success and failure are crisp and symmetric and accurately definable
- It should all be about 'positive' stories that make us feel good
- Method is context-free and universal; e.g. so-called “management theory”

... in justification, prediction and relative certainty.

Unpredictability and non-linear dynamics

- Periodic points are dense in the state space
  Contains possibly strong elements of regularity
- Topological transitivity
  Not decomposable into components
- Sensitivity to minuscule perturbation
  Future states are fundamentally unpredictable
- Possibly stochastically uncertain as well
- Need not be stationary or regular



(meta-)conjecture: the failure modes of ergodic models in non-ergodic environments are non-ergodic

- Unique transient states
- Deceptively long (O(log n)) seemingly predictable sequences
- More data does not give more information
- e.g. Deterministic K-systems & Bernoulli systems are strongly unpredictable

Unpredictability and incompleteness

"The mistake is thinking that there can be an antidote to uncertainty."

- Daniel Levithan

Fundamental uncertainty is underwritten by paradoxes, a paradox is a condensed infinite irreducible regress.

Prediction and observation are paradoxical, in general.

Incompleteness Phenomena

Provability logic has an implicit function theorem. Consider

$$f\ p \leftrightarrow \sim\Box p$$

The solution is a paradox, namely Gödel's second incompleteness theorem:

$$f = \sim\Box(\Diamond T)\rightarrow\Diamond T$$

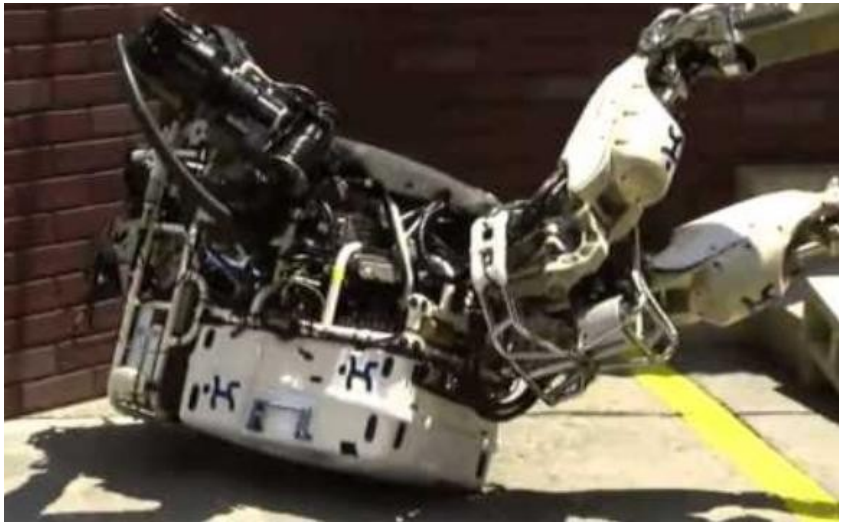
("If PA proves its own consistency, it is inconsistent; if PA is consistent, it cannot prove its own consistency")

Any such function* yields a fixed-point, a paradoxical generalisation of Gödel's Second Incompleteness Theorem. They are everywhere!

Autonomous dreams and automated realities

The number of operationally deployable autonomous systems worthy of the title is precisely zero.

It does not matter how good it is on average if we cannot withstand the consequences of its failure the first time its misguided expectations slam into nonconforming reality.



Convention*: plasticity* is the ability to socially cope with unpredicted and unpredictable future states of contested unstructured environments.

The machine is *autonomous* only to the extent that it manifests plasticity.

Revisiting military theory

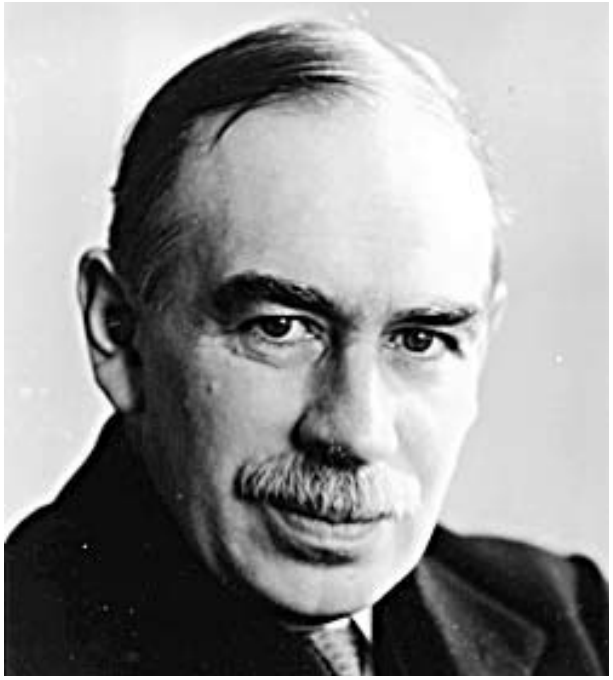
- Success is provisional and context-dependent
- Decisive failure is terminal: the operation is a failure, we are dead
- Normal failure is affordable and recoverable
- The justification programme forces small failures to accumulate under the surface and eventually explode into catastrophic failure
- Avoiding gross inefficiency and maximising efficiency are *not* the same
- 'Confirmation bias' is the basis of military deception

- Empirical knowledge is a tangled web of provisional ideas
- There is no such thing as a realisable universal positive method
- Ideas must be subject to strong responsible selection pressures
- Only what we can rule out is universal, giving an asymmetric demarcation

Autonomy as the allocation of scarce resources

How do agents in an economy behave under irreducible uncertainty?

- Assume existing conditions are a reliable guide
- Ignore what is unknown
- Rely on the prophesies of supposed experts.
- Assume these rituals are reliable despite the evidence.



They move capital between relatively certain but unprofitable liquid and uncertain but potentially profitable illiquid forms

- The result is economic volatility
- Small failures are allowed to accumulate into catastrophic failures
- Monetary policy – e.g. interest rates – moderates expectations

Bimodal investment portfolios

1) **Hedge against unacceptable outcomes**
   - Reason about sensitivity to failure
     - (maybe ergodic even though the overall problem is not?)
   - Expectation is that we have successfully hedged
     - → Seek refuting evidence
2) **Invest in opportunities that we can afford to lose**
   - Reason about affordable high potential returns
     - (maybe ergodic even though the overall problem is not?)
   - Expectation is that they will not pay off
     - → Seek refuting evidence

(A basis for an algorithmic view of economic information asymmetries)

On the utility of autonomy

The utility of autonomous systems lies in the potential to alter the of resource investment options

- The potential to reduce exposure to decisive failure
- Potential to allow greater investment in a wider range of affordable opportunities
- This means that unit costs of systems actually bear directly on their operational utility
- Economies of scale are operationally important, note the potential for economic absurdity with things that are too precious to use (conversely, what does *use* mean?)

Some technical enquiries

"Let go of certainty. The opposite isn't uncertainty. It's openness, curiosity and a willingness to embrace paradox, rather than choose up sides."

- Tony Schwartz

Non-stationary planning: non-classical logics (multi-modal & dynamic logics) and model generation for reasoning under uncertainty

Self-evolving Functions: Lambda calculi, Stochastic calculi, Quantum and quantum-typed calculi

Non-linear dynamical Systems: Ordinal structures, Entropy, Computational problems, Non-classical control

Simulation environments for autonomy

- Current simulations are high-fidelity and defined to eliminate the unexpected
- We need genuine unpredictability, at least from inside the game
- We want to discover, rather than merely script, behaviours (e.g. tactics)
- We could also discover, rather than merely script, scenarios
- Such simulations should be much better in finding sensitivity to failure
- The interactions between controllers in an environment creates the very problem controllers are there to solve
- unpredictable future states as described by Clausewitz
- more data will not normally give more information
- unique transient states – punctuated equilibria – and their disintegration are the scenarios

Fidelity should be traded for unpredictability, not the other way around.

In closing

"But in war, as in life generally, all parts of the whole are interlocked and thus the effects produced, however small their cause, must influence all subsequent military operations and modify their final outcome."

- Karl von Clausewitz

UNCLASSIFIED

Australian Government
Department of Defence
Defence Science and Technology Group

Unpredictable Outcomes in Unstructured Environments

the future of machine reasoning

Dr Darryn Reid
Principal Scientist
Joint & Operations Analysis Division
Defence Science & Technology Group

DST GROUP | Science and Technology for Safeguarding Australia

The third choice of the military enlightenment

"Though our intellect always longs for clarity and certainty, our nature often finds uncertainty fascinating."

"The very nature of interaction is bound to make it unpredictable."

"The difficulties accumulate and end by producing a kind of friction that is inconceivable unless one has experienced war."

"Circumstances vary so enormously in war, and are so indefinable, that a vast array of factors has to be appreciated."

- Karl von Clausewitz

I will introduce the military enlightenment and its three choices: the first in favour of rationality, the second in favour of the sciences for rational method, and the third, which was the misstep, in favour of linear convergences, relative certainty, and knowledge as accumulation of facts. One prominent thinker stood out, pillorying his contemporaries for embracing the clockwork universe and the inductivist model of science that Bacon formulated by secularising pre-religious mysticism. He was Carl von Clausewitz. Mind you, plenty of other people through history also rejected justification, prediction and relative certainty, including Galileo, and notably the first identifiable scientist, Thales. The major point here is that knowledge is not smoothly convergent, but highly non-linear: critical, creative and inherently disruptive.

I want to particularly emphasise Clausewitz's characterisation of unpredictability as the result of interaction, from friction, and unpredictability from stochastic chance. In a very deep sense, he foreshadows the basic model of modern non-linear dynamics and complexity theory around 150 years before the mathematics was known. This contrasts sharply with the linear certainty-oriented views of his contemporaries, and the divide continues more or less to this day. Belief in certainty is the basis for military deception – and all war is based on deception.

Beliefs in certainty show up in the reduction of complex issues to simple linear formulaic absolutes. These have emerged and re-emerged constantly throughout history, of course: in modern times we have seen cults of certainty in military theory, at least twice in economics, business administration management theory with its claims of universality of its supposed methods, to name a few examples. In military theory, such ideas tent to strongly hinge on techno-worship, pompous grandiosity, and assertions of some supposedly complete truth or another. Bridging the taboo of crossing between military theory and politics is necessary if we want to really understand these phenomena - certainty cults in military thinking are invariably tied to and reflect the political ideologies of the day.

So much for dismissing seemingly esoteric 'high theory', from which we too often want to distance ourselves, in favour of supposed 'doing practical things'. We prefer to hide in our little hermetically sealed boxes of specialist expertise. Thus we marginalise memory, imagination, intuition, ethics and cling instead to a narrow self-serving parody of the quality of analytical reason. All these qualities, and possibly others, are necessary to rationality, the genuine article, and thus what we have too often been left with is a non-rational parody of rationality.

Clausewitz is often quoted in the context of simple absolute truths, though Clausewitz himself was way too sophisticated to topple into ideological exactitude, and indeed warned against it. For instance, I have seen cults of military thinking come and go that worship technology for its own sake, that see science as merely the means of technology production, which actually stands in open contradiction of Clausewitz. This is the vile reduction of science down to mere technical mastery. As another example, we reduce power to linear measures of numbers of men and materiel, again in open contradiction of Clausewitz and Sun Tzu, for example, who cast power in distinctly non-linear terms. The vestige of our conceit is our marginalisation of and thus continued inability to really deal effectively with so-called asymmetric opponents.

Such self-contradiction is by no means rare. Our ideas are often frozen at the level of infantile linear absolutes. The ideologies so enamoured with technology find themselves in a state of gross inconsistency with the founding science behind the very technologies they worship. My first point of business is thus two-fold: autonomy does not lack for excited futurologists who are infatuated with technology and high on grand historical predictions, nor for practitioners too embedded in small worlds of specialist expertise to pop their heads above the parapet and actually examine critically their foundational assumptions against the reality of the world in which these things are supposed to do jobs for us.

UNCLASSIFIED

The mechanistic paradox

People tend to believe

- Knowledge is gained by 'objective' observers gorging on data and extracting the truth out of it
- Every event has a necessary and sufficient set of causes
- Every question has a complete and correct answer, and that it is obtainable, at least in principle
- Decisive action hinges on relative certainty
- Future outcomes can be predicted by governing laws and a sufficiently accurate account of the past
- All failure is inherently pathological
- More data gives more information
- Incremental improvements will eventually accumulate into success

... we live in a clockwork managerial universe.

I have spent many years now writing and presenting about epistemology, methodology, meta-mathematics, mathematics and computer science in relation to the nature of knowledge and the means by which we can obtain it. This represents something of a quick overview of the undercurrent beneath commonly held but logically unsustainable beliefs about knowledge as the search for, attainment of, or at least convergence towards, certainty.

The first point is about the widespread belief in inductivism – I've published elsewhere pointing out that is actually secularised pre-religious mysticism – which holds that knowledge is the result of inductive generalisation approaching the truth from accumulation of facts. I have it here because it is the assumption behind typical bottom-up development, and this explains why bottom-up development invariably fizzles out and fails.

The second concerns the belief that every event has necessary and sufficient causes; whenever we asked why something happened, we can obtain plausible answers in terms of chains of supposed causative prior events. Yet the deeper we dig into this, the more supposed causes actually look like epistemic factors that influence our decisions – reasons for acting a certain way – and the less they look like immutable ontological features of the environment.

Causal determinism is really about predictability, and has nothing to do with determinism, which is about the absence of arbitrary choice in the application of transition operators. Deterministic systems can be unpredictable and non-deterministic systems can be predictable. Not only are their events that do not have causes, but the inference from sets of necessary and sufficient causes to predictability simply doesn't follow either. To refute the belief, I don't need to assert that no events ever have causes, nor even reject the assertion that every event has causes, merely that causes may be necessary but not sufficient for some events some of the time.

People tend to believe that all failure is pathological because they believe that we can and therefore should be correct.

I have often encountered discomfiture with saying that these beliefs are false, on the basis that uncertainty and the limits of knowledge makes people feel bad. I find this deeply troubling: the justification programme is at least latently horrific: it is inherently absolutist, with totalitarian and authoritarian ambitions never far away. There is very dark psychology at work here to do with intolerance of ambiguity and uncertainty, and social domination orientation. But this is for another time.

Methodological hallucinations

People also tend to believe

- We can precisely define our problem situations
- We can accurately define complete and correct solutions

- The outcomes are pre-determinable
- The path from problems to solutions is a linear, a matter of efficiency and expected utility
- Success and failure are crisp and symmetric and accurately definable
- It should all be about 'positive' stories that make us feel good
- Method is context-free and universal; e.g. so-called “management theory”

... in justification, prediction and relative certainty.

I might not talk too much about these beliefs in that setting of management so-called theory that I do so love to hate. The beliefs discussed here are not unique to management theory, yet management theory does provide a highly topical example, and one that has enormous influence to our ongoing disadvantage. I will avoid the temptation to go completely through the history of management theory and its flaws from its inception that leads me to rank it for legitimacy of its claims to scientific status just below those of Scientology, in the interests of time.

The points I really want to emphasise here is the linearity of the alleged means of producing solutions to problems, the foolishness of assuming that problems and solutions are crisp, that goals can be known precisely in advance, and especially that people tend to assume that success and failure are symmetric. In other words, anything that is not a success is a failure, and vice-versa. Like it or not, it is a reasonable description of both how we mainly approach autonomous systems development, and of the expectations that the current ostensibly autonomous systems take into their environments.

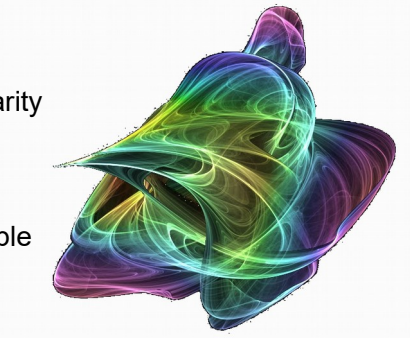
We observe the universality assumption in management theory, which openly holds that managerial methods are independent of context. The same claims punctuated military theory in terms of the Third Choice. This openly promoted as scientific, but is merely pseudo-science of the most disingenuous kind. Consequently, we have suffer the absurd results from the absurd consultancy practices of obtaining advice about how to structure our businesses from people who know nothing at all about our business.

To be fair, science has for a long time embodied and promoted the idea that certainty is obtainable, at least in principle, and mathematics has often been seen - and promoted - to constitute the pinnacle of this programme. Perhaps it is no wonder that revealing the true nature of mathematics to those not familiar with the mathematics of this is so often met with incredulity and silly statements about optimism and the like. This probably captures a pretty decent summary of public perceptions about science (or certainly that promoted by marketing, at least). Yet it is a distinctly false view. Science is about obtaining reliable knowledge, which is fundamentally not the same thing as obtaining certainty.

UNCLASSIFIED

Unpredictability and non-linear dynamics

- Periodic points are dense in the state space
  - Contains possibly strong elements of regularity
- Topological transitivity
  - Not decomposable into components
- Sensitivity to minuscule perturbation
  - Future states are fundamentally unpredictable
- Possibly stochastically uncertain as well
- Need not be stationary or regular



(meta-)conjecture: the failure modes of ergodic models in non-ergodic environments are non-ergodic

- Unique transient states
- Deceptively long (O(log n)) seemingly predictable sequences
- More data does not give more information
- e.g. Deterministic K-systems & Bernoulli systems are strongly unpredictable

There are stronger and weaker definitions of chaos, but this one best serves my purposes here. The presence of regularity is highly seductive: it may be even under complete chaos that a system will contain a remarkably strong component of regularity and yet be fundamentally unpredictable. We tend to forget about the second point as well as the third, and assume we can decompose for our analytical pleasures systems that cannot be decomposable. We can only do this if we smooth away the unpredictable component. We are smoothing away the very things that matter most, just so we can use our favourite methods that appear to work for us. For a while. The final point in the top box is about what happens if we introduce stochastic uncertainty as well, to produce stochastic non-linear systems. The combination of even well-behaved distributions with non-linear perturbations does not produce well-behaved distributions.

Notice systems manifesting very strong forms of uncertainty, are still fully deterministic. Determinism is often misunderstood and misused as a concept – non-determinism and indeterminism are very different things. Non-determinism is about the presence of arbitrary choice in the application of transition operators; causal determinism is a position not about determinism but is an argument (with a premise and a conclusion) about predictability: it asserts that every event has a necessary and sufficient set of causes, and concludes on this basis that the world is predictable given sufficient knowledge of the supposed governing laws and of past history. The premise simply does not support the conclusion: even completely deterministic mechanistic systems are unpredictable in general. Moreover, the premise is false as well.

Returning to the commonplace belief that more data provides more information about a system: suppose we ask about exactly what conditions we have to have in place for this to be true. It turns out that the conditions required to make this true are amazingly restrictive. The system has to be stationary, meaning that the generators never change, and it has to be regular, meaning that the generators meet an extraordinarily high standard of polite behaviour. The terribly convenient consequence is that long realisations of the process – long sequences of sampling – converge to the ensemble average. I'm unsure as to what I find more disturbing: that people nonetheless believe that more data gives us more information in general, or that almost all projects to build supposedly autonomous systems rely on some form of ergodicity.

What happens when we break the ergodicity assumption? More data does not give more information about the system, and instead the system manifests unique transient states. Different kinds of systems that are not ergodic, such as K-systems and Bernoulli systems, display very strong forms of unpredictability; yet even here, we are still assuming we have complete sample sets over all possible outcomes, so it gets even more interesting again when we start to talk about handling problems in the real world where we cannot expect to have sample sets at all.

Unpredictability and incompleteness

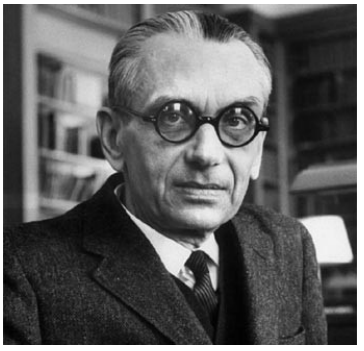
"The mistake is thinking that there can be an antidote to uncertainty."

- Daniel Levithan

Fundamental uncertainty is underwritten by paradoxes, a paradox is a condensed infinite irreducible regress.

Prediction and observation are paradoxical, in general.

Incompleteness Phenomena



Provability logic has an implicit function theorem. Consider

$f\ p \leftrightarrow \sim\Box p$

The solution is a paradox, namely Gödel's second incompleteness theorem:

$f = \sim\Box(\Diamond T)\rightarrow\Diamond T$

("If PA proves its own consistency, it is inconsistent; if PA is consistent, it cannot prove its own consistency")

Any such function* yields a fixed-point, a paradoxical generalisation of Gödel's Second Incompleteness Theorem. They are everywhere!

I will discuss a little about the mathematics of irreducible uncertainty, touching on non-linear dynamics, and connecting this to a computational view through partitioning the system into cells, and then I will talk about incompleteness illustrating with Gödel's second incompleteness theorem. Uncertainty and unpredictability boils down to logical paradox, for an obvious instance think of the self-reference of the self-fulfilling (Oedipus effect) and the self-defeating prophesy. Incompleteness phenomena are really all about self-referential statements and self-referential statements mean, in general, lots of paradoxes; paradoxes are, in turn, really condensed infinite regressions. I also might talk a little about why, far from being pessimistic, these *facts* of existence in this universe are natural, and positive; I am disturbed when I hear people suggesting that the hope of somehow overcoming such fundamental limits on knowledge is optimistic. No, such a thing is merely convenient in a narrow utilitarian sense, betrays a lack of understanding about incompleteness phenomena, and a lack of understanding that the obsolete totalitarian absolutist position that such a hope represents is the darkest of all places.

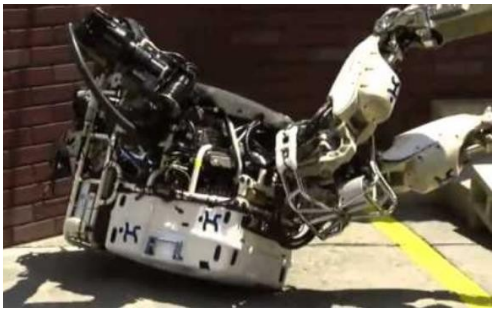
I might elaborate a little on the logic about uncertainty as logical paradox, which is a natural result of self-reference, and paradox is a condensed form of infinite regress, a real looking-glass into the infinite complexity of all of mathematics. Gödel's famed First Incompleteness Theorem starts with the logical system built around Peano Axioms for numbers. An important point here is that anything can be encoded as numbers, including statements in this logic about numbers. Using a clever Quining technique, Gödel wrote a statement that makes claims about its own encoding as a number, and hence about itself, specifically asserting its own unprovability (but not about its own truth!). I'm a simple man with a small brain, so I enjoy playing with this kind of thing using approaches such as a modal logic called Provability Logic more than I do in Peano Logic itself. Provability Logic is propositional logic with a modal operator $\Box$ "provable" and its dual $\Diamond$, and Löb's theorem: $\Box(\Box a \rightarrow a) \rightarrow \Box a$; the second incompleteness theorem can be written in Provability Logic as $\sim\Box(\Diamond T) \rightarrow \Diamond T$, which means that if Peano Logic can prove its own consistency then it is inconsistent, or if it is consistent, it cannot prove its own consistency. Our Provability Logic system comes with an implicit function theorem, which means we can write self-referencing equations and provided they are 'modalising' then they have solutions that do not involve any self-reference. Self-reference is less fundamental than we sometimes think – though unfolding self-reference in general gives infinite results. Turing's Halting problem looks exactly like this: the Halting Problem is actually a paradox, and its original proof it is an infinite regress that essentially defines an uncomputable number.

I hope to have made the point that paradoxes arise everywhere, we are swimming in them all the time, and that their nature and consequences are not some strange pathological phenomenon confined to the dark recesses of pure mathematics and theoretical computer science, hidden away from a clockwork reality. Uncertainty and unpredictability and incompleteness are natural and normal, while it is predictability and relative certainty that are strange and unusual and pathological. Isn't it strange how people, self-purporting to strains of practicality, who used to dismiss the consequences of theory as mere pure ideas divorced from the messy real world, turn around after encountering mathematical incompleteness and its implications for their precious beliefs in the attainment of certainty by linear managerial processes, dismiss the incompleteness phenomena of abstract theory as being too messy for their allegedly tidy and predictable 'real' world?

Autonomous dreams and automated realities

The number of operationally deployable autonomous systems worthy of the title is precisely zero.

It does not matter how good it is on average if we cannot withstand the consequences of its failure the first time its misguided expectations slam into nonconforming reality.



Convention*: plasticity* is the ability to socially cope with unpredicted and unpredictable future states of contested unstructured environments.

The machine is *autonomous* only to the extent that it manifests plasticity.

I will talk about the gulf that lies between the things we have – which I propose are not autonomous but merely automations – and the kind of autonomy we require, and explain the distinction in terms of the inability for our current methods to deal effectively with the kinds of fundamental uncertainties intrinsic to interacting with hostile and non-hostile elements in unstructured environments, the kinds of profound uncertainties recognised by Clausewitz (and Sun Tzu, for that matter).

The promise of autonomous systems and the reality of the automations we have are two very different things; you only have to see what supposedly autonomous systems *cannot* do, and how much human intervention and control they require to do what they can do to realise that the gap is enormous. I cannot recall how many times I've seen someone claiming – usually tacitly, though sometimes explicitly – that more of what we have been doing so far will solve the problem. Funding the development of a better electronic eyeball, for instance, may very well be a worthwhile thing to do, but to suggest that by doing so we will bring into existence the autonomous systems of our vertiginous dreams is, frankly, ludicrous.

That many commercial enterprises manifest ostensibly autonomous systems begs many to question why Defence cannot make the same boast. Yet a better question would be whether these marvels are indeed autonomous. Defence represents an especially challenging set of circumstances: we do not have the luxury of tightly controlling our environments so that the strong expectations built into our systems will not crash unhappily into reality. Moreover, for us, the cost of failure is especially high.

A core theme of this presentation is that bottom-up development, oriented around successively accumulating cases under strongly predictive models of control, can never yield an aggregation that suffices to produce operationally usable autonomous systems. There are fundamental barriers that cannot be breeched from within the sets of assumptions on which bottom-up development pivots.

So the question that the rest of the presentation will address is simply “Where to from here?”. The answer, which is the result of some of my personal research over the last 18 months or so, has defined the strategic research direction of DST Group's Project Tyche, and is apparently diffusing out through our strategic relationships to academia and parts of industry as well. I do not claim that this is correct and complete, but rather submit that the approach I will promote today is a little less wrong than those I intend it to displace.

Revisiting military theory

- Success is provisional and context-dependent
- Decisive failure is terminal: the operation is a failure, we are dead
- Normal failure is affordable and recoverable
- The justification programme forces small failures to accumulate under the surface and eventually explode into catastrophic failure
- Avoiding gross inefficiency and maximising efficiency are *not* the same
- 'Confirmation bias' is the basis of military deception

- Empirical knowledge is a tangled web of provisional ideas
- There is no such thing as a realisable universal positive method
- Ideas must be subject to strong responsible selection pressures
- Only what we can rule out is universal, giving an asymmetric demarcation

I will talk about Keynes' insight into exactly the kind of uncertainty that Clausewitz talked about, and that arises in the formal settings of the fields of what I described earlier as post-modern mathematics. Ironically, economics has been tarnished with some of worst examples of ideological belief in the attainability of certainty, assumptions of predictability, and subsequent linear oversimplification masquerading as science, yet also holds some of the richest insights into the nature of uncertainty and especially when it comes to considering how to deal with it effectively. So it mirrors military theory in its schizophrenia.

Agents in an economy face an essential problem: they must make investment decisions that will play out in a future they cannot control nor really predict. Keynes argued that the future is fundamentally uncertain – not merely stochastically random – because the sample sets of possible outcomes cannot be determined in advance, if at all. Keynes defines investment as the allocation of resources under irreducible uncertainty – the same uncertainty as interested Clausewitz, and that punctuates fundamental mathematics and computer science. To cope, our economic agents resort to superstitious rituals; Keynes identified four, which I have summarised on the slide. In short, they retreat to the justification and relative certainty programme, to clockwork managerialism. The consequence is that expectations are extremely fragile and economies are highly volatile, with bubbles and bursts that are wholly generated from within the system (just as in nonlinear dynamics).

Keynes also described how our agents, in response to doubt about future returns, hold more liquid assets, which are relatively safe but unprofitable. The store of wealth in liquid form is a measurable manifestation of our agents' confidence in future returns. Lower confidence requires higher interest rates to entice them to draw capital out of liquid form and place it into profitable but volatile illiquid assets – especially production assets. In stark contrast to neoclassical laissez fare, Keynes asserted a role for government intervention through monetary policy to moderate the boom and bust of outrageous market circumstance.

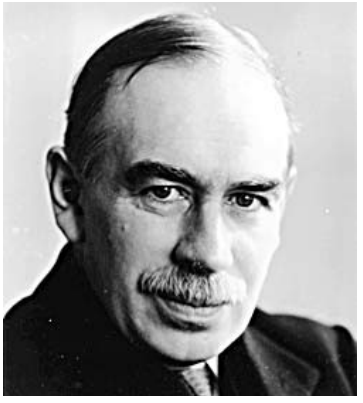
What Keynes did not ask – and I am asking now – is what happens instead if our agents do not resort to the superstitious rituals of the positive justification and relative certainty programme, and instead face up to the complex, asymmetric nature of knowledge. They face a fundamental tension between holding their limited assets in slowly degrading liquid form and placing them in potentially profitable but volatile space of illiquid investment opportunities. They face a bimodal tension, with countable sets of possible hedging options and countable sets of possible opportunities. Noting that these sets may be infinite in theory and that not all options may be visible – thus we will have computational complexity questions dropping out of this all over the place, which makes me very happy.

UNCLASSIFIED

Autonomy as the allocation of scarce resources

How do agents in an economy behave under irreducible uncertainty?

- Assume existing conditions are a reliable guide
- Ignore what is unknown
- Rely on the prophesies of supposed experts.
- Assume these rituals are reliable despite the evidence.



They move capital between relatively certain but unprofitable liquid and uncertain but potentially profitable illiquid forms

- The result is economic volatility
- Small failures are allowed to accumulate into catastrophic failures
- Monetary policy – e.g. interest rates – moderates expectations

Suppose we have our agents face up to the bimodal nature of their situation, as a means of actually dealing with the fundamental uncertainty they face. The picture represents their circumstances: the Gaussian in the middle is a fiction, and the practice of trying to ride its middle – the mean and variance are also shown – is a fiction that will eventually fail non-ergodically. Their reality is the blue distributions, which are non-stationary and arbitrarily shaped, and are fundamentally unknowable. Our agents cannot even obtain meaningful sample sets, so the axis is itself opaque.

The constitutive point is to create a favourable asymmetry, with the implication that they forego the temptation to try to get rich quickly by riding the bubble, the middle; they are now in the game to survive and prosper over the long term. Our new agents will first utilise their available resources to hedge against unacceptable outcomes – this means that our agents need to be able to effectively decide what constitutes an unacceptable outcome in the present circumstances. Having done this, they may then invest what they can afford to lose in opportunities that will only sporadically and unpredictably pay returns, but will the reward will be large when they do so. This requires that they need to be able to decide affordable opportunities. In neither case are they trying to predict future outcomes, only determining their sensitivity to failure and determining what constitutes a worthwhile opportunity, which is, as far as I know, a new basis for a novel algorithmic construction for economics concerning information asymmetries (Stiglitz, Krugman and others). As an aside – I've also started a little project for a quantum algorithmic theory of economic information based on my idea about this; the motivation is to look at economic entropy as irreversibility, irreversibility in economic systems as information asymmetry.

There are a few things to note. Firstly, our new agents are inherently taking a logically negative stance, oriented around refutation. On the tail of unacceptable outcomes, the standing theory is that our agent has successfully hedged, and it looks for conflicting information that warns of impending refutation of this theory. Evidence to supposedly 'support' the theory that the unacceptable is successfully hedged is *irrelevant*. On the upper tail of opportunity rewards, the standing theory is that the agent's opportunity investment will go bust, and what attracts the agent's attention is the refutation of this theory, namely that an opportunity bet is paying off. Having taken an opportunity bet, evidence that it is failing is *irrelevant.*

Bimodal investment portfolios

1) Hedge against unacceptable outcomes
- Reason about sensitivity to failure
  - (maybe ergodic even though the overall problem is not?)
- Expectation is that we have successfully hedged
  - → Seek refuting evidence

2) Invest in opportunities that we can afford to lose
- Reason about affordable high potential returns
  - (maybe ergodic even though the overall problem is not?)
- Expectation is that they will not pay off
  - → Seek refuting evidence

(A basis for an algorithmic view of economic information asymmetries)

I will talk about the bimodal nature of a rational investment strategy in the face of unpredictable future outcomes. This will lead into my criticism of current objective function choices, the strength of their underpinning assumptions about the nature of the environment, and the largely unacknowledged depth to which they penetrate our current methods. My intention is that this economically inspired view of autonomous systems as resource allocation leads to the definition of whole new kinds of objective functions – the central point here is that this inspires new objective functions squarely aimed at hedging against unacceptable outcomes, at least, if not also making affordable opportunity bets as well.

Let's have our agents face up to the bimodal nature of their situation, as a means of actually dealing with the fundamental uncertainty they face. The picture represents their circumstances: the Gaussian in the middle is a fiction, and the practice of trying to ride its middle – the mean and variance are also shown – is a fiction that will eventually fail non-ergodically. Their reality is the blue distributions, which are non-stationary and arbitrarily shaped, and are fundamentally unknowable. Our agents cannot even obtain meaningful sample sets, so the axis is itself opaque.

The constitutive point is to create a favourable asymmetry, with the implication that they forego the temptation to try to get rich quickly by riding the bubble, the middle; they are now in the game to survive and prosper over the long term. Our new agents will first utilise their available resources to hedge against unacceptable outcomes – this means that our agents need to be able to effectively decide what constitutes an unacceptable outcome in the present circumstances. Having done this, they may then invest what they can afford to lose in opportunities that will only sporadically and unpredictably pay returns, but will the reward will be large when they do so. This requires that they need to be able to decide affordable opportunities. In neither case are they trying to predict future outcomes, only determining their sensitivity to failure and determining what constitutes a worthwhile opportunity, which is, as far as I know, a new basis for a novel algorithmic construction for economics concerning information asymmetries (Stiglitz, Krugman and others). As an aside – I've also started a little project for a quantum algorithmic theory of economic information based on my idea about this.

There are a few things to note. Firstly, our new agents are inherently taking a logically negative stance, oriented around refutation. On the tail of unacceptable outcomes, the standing theory is that our agent has successfully hedged, and it looks for conflicting information that warns of impending refutation of this theory. Evidence to supposedly 'support' the theory that the unacceptable is successfully hedged is *irrelevant*. On the upper tail of opportunity rewards, the standing theory is that the agent's opportunity investment will go bust, and what attracts the agent's attention is the refutation of this theory, namely that an opportunity bet is paying off. Having taken an opportunity bet, evidence that it is failing is *irrelevant.*

On the utility of autonomy

The utility of autonomous systems lies in the potential to alter the of resource investment options

- The potential to reduce exposure to decisive failure
- Potential to allow greater investment in a wider range of affordable opportunities
- This means that unit costs of systems actually bear directly on their operational utility
- Economies of scale are operationally important, note the potential for economic absurdity with things that are too precious to use (conversely, what does *use* mean?)

Here I am viewing military command and control in an operational setting – and I regard the decision-making of our commercial cousins as being precisely the same fundamental problem, albeit wrapped in different terminological constructs – as a matter of allocating scarce resources to possible tasks. The machine has to be able to manage its resources under uncertainty, and, in turn, the system that uses the machine has to be able to do the same thing – the machine is, in turn, itself a resource. Again, notice the kind of cascading self-reference of the construction!

This brings us to a fundamental question that has not, to my knowledge, been previously been addressed in a adequate manner: what is the utility of autonomous systems? “Dirty, dull and dangerous” simply does not cut the mustard. Taking this resource allocation viewpoint, I conclude that the utility lies precisely in the potential for the autonomous system to favourably alter the balance of investment options faced by the operational commander. There are two ways in which this can happen, which we already know about from the preceding economic view of decision-making under uncertainty. The first is by allowing the commander new means for hedging against decisive failure, and the second is by permitting the possible investment of resources into opportunities; previously unaffordable failures are now potentially rendered affordable with the cost of failure being the loss of a machine rather than the loss of human life, for instance.

This economic view reveals the startling new conclusion that the cost of acquisition, maintenance and application of autonomous systems bears directly on the operational utility of the system. This draws into question where the boundaries lie – remember that they are rubbery and context-dependent – regarding the cost-benefit trade-off of such systems. Arguably, some systems in Defence are already pushing these boundaries of absurdity, though we also have to be careful: such systems may be more effective at altering the range of investment options at a strategic or operational level than at a tactical one.

Some technical enquiries

"Let go of certainty. The opposite isn't uncertainty. It's openness, curiosity and a willingness to embrace paradox, rather than choose up sides."

- Tony Schwartz

Non-stationary planning: non-classical logics (multi-modal & dynamic logics) and model generation for reasoning under uncertainty

Self-evolving Functions: Lambda calculi, Stochastic calculi, Quantum and quantum-typed calculi

Non-linear dynamical Systems: Ordinal structures, Entropy, Computational problems, Non-classical control

This is about one of my theoretically oriented efforts to derive an algorithm from my bimodal strategy. The ability to handle unpredictable events in general is the uncomputable universal autonomy that we might like in principle. The ability to handle unpredicted events is the computable autonomy that will suffice in practice. The development of the former in theory is the foundation for realising the latter in practice; in the middle sits the computer science underpinning the caveat on the convention stating "...within bounds we consider operationally acceptable". I envisage here not just the kinds of algorithmic complexity resource constraints we are used to in the theoretical computer science literature, but also the problem sub-classing that captures the context dependence, situation sensitivity and embodiment we already intrinsically understand is vital to engineering workable solutions.

I'm attracted to the high-level semantics of Lambda Calculi, though the non-primitive nature of reduction is inconvenient when considering computational resource use. I have some of the ingredients in place for what I envisage, using encodings of Lambda Theory inside Lambda theory itself and absolute complexity to create self-modifying programs. The key, I think, to making a workable self-modifying program is to use semantic rather than naïve notions of resource bounding – this is the "within acceptable limits" part of the convention. I am hoping to use this to produce small implementable self-modifying abstract machines in the near future. I don't imagine such things to be anything so grand as the final word in artificial intelligence, but rather as a kind of toy mathematical machinery implementing my bimodal strategy by which to advance the growth of knowledge by a little bit.

My final note is an observation about mathematics and theoretical computer science, and indeed about science in general. It is my response to the beautiful grand unknowability of things; kind of a flip-side to the dark and stupid mechanistic interpretation of mathematics as a provider of absolute certainty. Call it a consequence of incompleteness, and it should be read as an optimistic statement, not a negative one.

UNCLASSIFIED

Simulation environments for autonomy

- Current simulations are high-fidelity and defined to eliminate the unexpected
- We need genuine unpredictability, at least from inside the game
- We want to discover, rather than merely script, behaviours (e.g. tactics)
- We could also discover, rather than merely script, scenarios
- Such simulations should be much better in finding sensitivity to failure
- The interactions between controllers in an environment creates the very problem controllers are there to solve
- unpredictable future states as described by Clausewitz
- more data will not normally give more information
- unique transient states – punctuated equilibria – and their disintegration are the scenarios

Fidelity should be traded for unpredictability, not the other way around.

This is picking up the point raised earlier about the typical practice of assuming a clean separation between problem and solution. Simulation and modelling usually follow this practice; the question that is not always seriously considered is whether the assumption that the phenomena of interest are not topologically transitive is actually reasonable. It may be so in many applied problems, but it certainly is not in many others.

A project I'm involved with began with the desire build a simulation environment – really a computational model – in which we might discover tactics, rather than script them in and then collect statistics – smoothing away the supposedly rare events on the tails, of course – about their efficacy.

My counter-question concerns how to go a step further, and discover scenarios as well. We have already discussed all the ingredients needed to do this: the controllers themselves generate the problem that they are there to solve, and their interactions generates unpredictability. We could measure their ability to do this through order parameters on the resulting overall model – I am interested in algorithmic complexity measures here, as well as entropy, especially entropy as irreversibility; in other words, entropy as asymmetry in the resulting dynamics.

Where do our scenarios come from? They are precisely the unique transient states that non-ergodic systems such as K-systems or Bernoulli systems will necessarily manifest.

This approach directly runs counter to the typical manner of simulation systems that I have witnessed: acknowledging that resources are limited, success pivots on unpredictability at the expense of fidelity of representation rather than fidelity at the expense of unpredictability. In other words, unpredictability is desirable.

Note that this does not somehow suggest that we do intend to have models. It means we are trying to change exactly what we choose to model.

In closing

"But in war, as in life generally, all parts of the whole are interlocked and thus the effects produced, however small their cause, must influence all subsequent military operations and modify their final outcome."

- Karl von Clausewitz

I will close by referring back to Clausewitz for the last word on the central importance for all defence capability to hinge on the ability to handle an irreducibly uncertain operating environment.

This is really just a summary of some of the main points: we cannot expect to solve a problem that we don't acknowledge, and the first point of business in my role with respect to autonomous systems as Principal Scientist is to garner broad acknowledgement of the problem. Then it's on to more interesting things: I propose that the fundamental reason for the problem is the inability of current systems (and for that matter, a whole lot of modelling outside of autonomous systems research as well) to adequately handle unpredicted future states of the intended operating environment.

Why is the case? Because of widespread beliefs in certainty and justification, which is not helped by modern management theory, and – frankly – was not helped either by the stupid linear account of science and of mathematics that held that science and mathematics is the provider of absolute certainty or at least that it converges monotonically towards it. This was the view of Descartes, and it still holds sway in the public today. What can be said for a man who spent most of his life trying to separate his mind from his body?

I propose a new logically negatively oriented convention defining autonomy to supersede the many positively oriented conventions around today – they fail predictably because of their positive orientation. Though I've tried to couch it in acceptable terms, just under the surface, it is more about ruling out what is not autonomous than it is about trying to define what is autonomous. I am against notions of positive universal method for strong mathematical and philosophical reasons, and the fact that a logically negative position like this annoys people by asking them to think is not going to get in the way. Call it a minimum concession to the realities of context-sensitivity, embodiment, resource limitations and creativity – both human and machine – that rules out unacceptable research and development problem choices.