A Multidisciplinary Approach to Complex Systems Design

The University of Adelaide
Discipline of Applied Mathematics

Alex Ryan
B.Sc (Applied Mathematics) (Hons)

Thesis submitted for the degree of Doctor of Philosophy
March, 2007

To Eleanor, for ensuring this thesis did not consume me.

I may not have gone where I intended to go, but I think I have ended up where I intended to be.

Douglas Adams

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis being made available in the University Library.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder/s of those works.

A. J. Ryan and A. Grisogono. Hybrid Complex Adaptive Engineered Systems: A Case Study in Defence. *Interjournal*, 2004.

M. Prokopenko, F. Boschetti, and A. J. Ryan. An Information-theoretic Primer on Complexity, Self-organisation and Emergence. *Submitted to Complexity*, 2007.

A. J. Ryan. Emergence is coupled to scope, not level. *Submitted to Complexity*, 2007.

A. J. Ryan. About the bears and the bees: Adaptive approaches to asymmetric warfare. *Interjournal*, 2006.

A. J. Ryan and D. O. Norman. Agile and Flexible Capability Development. In *Proceedings of the Land Warfare Conference*, Brisbane, Australia, 2006.

Signed:

Date: June 18, 2007

Acknowledgements

I have been most fortunate to be embedded in three stimulating research environments during my PhD program. The Land Operations Division of DSTO has provided outstanding support for my study, showing an admirable degree of trust and flexibility towards the research threads I have pursued during my candidature. My DSTO supervisor Darryn Reid has been an inspiration due to his willingness to challenge widely held beliefs; and the clarity of thought and argument demanded of others but especially himself. I blame Darryn for many rewarding diversions into philosophy.

My Research Leader Anne-Marie Grisogono has been most influential in persuading me of the importance of adaptivity for coping with complexity, and could not have more enthusiastically championed our research in this area. The thirteen members of my Complex Adaptive Systems Long Range Research task over the previous three years have been willing accomplices in progressing beyond the buzzwords and identifying the value of complex systems research for defence. Scott Wheeler, Martin Burke and David Matthews from the Defence Systems Analysis Division of DSTO also deserve special mention for their encouragement and continued engagement with my research.

My University of Adelaide supervisor Nigel Bean has gently steered me towards the completion of this thesis. I have appreciated Nigel's emphasis on real world impact, and his genuine and selfless interest in my development. The Teletraffic Research Centre at the University of Adelaide, popularly led by Mike Rumsewitz, has always been welcoming and provided me with a home that seemed to be far too much fun to be as productive as it was.

The third environment I wish to acknowledge is the CSIRO Complex Systems Science Theme. John Finnigan has built the leading Australian complex systems community of practice by fostering an exciting, open and friendly research culture that has generated exceptionally high levels of scientific excellence, collaboration and morale. CSIRO has always welcomed my involvement in their workshops and interaction tasks, which have enriched my understanding of emergence, self-organisation and complexity and been a source of mentors, peers and coauthors for whom I have the greatest respect.

During my candidature two courses were influential: Introduction to Complex Systems taught by Yaneer Bar-Yam and Cognitive Science taught by Gerard O'Brien and Jon Opie. I am sure they would each recognise their influence in this document, and I thank them for challenging my theories and sharing some of their own.

Lastly and perhaps most importantly I would like to thank my coauthors, whose ideas have interwoven with my own to create a whole that is surely greater than the sum of its parts. The joint papers that have been published during my candidature include:

[24] T. J. A. Baker, M. Botting, M. J. Berryman, A. J. Ryan, A. Grisogono, and D. Abbott. Adaptive Battle Agents: Emergence in Artificial Life Combat Models. In *SPIE: Smart Structures, Devices, and Systems*, 2004.

[139] A. Grisogono and A. J. Ryan. Designing Complex Adaptive Systems for Defence. In S. Wilson, editor, *System Engineering Test and Evaluation Conference*, Canberra, Australia, 2003.

[209] K. L. Lim, I. Mann, R. Santos, B. Tobin, M. J. Berryman, A. J. Ryan, and D. Abbott. Adaptive Battle Agents: Complex Adaptive Combat Models. *Proceedings SPIE*, 6039:48–60, 2005.

[252] M. Prokopenko, F. Boschetti, and A. J. Ryan. An Information-theoretic Primer on Complexity, Self-organisation and Emergence. *Submitted to Complexity*, 2007.

[269] A. J. Ryan and A. Grisogono. Hybrid Complex Adaptive Engineered Systems: A Case Study in Defence. *Interjournal*, 2004.

[270] A. J. Ryan and D. O. Norman. Agile and Flexible Capability Development. In *Proceedings of the Land Warfare Conference*, Brisbane, Australia, 2006.

[271] A. J. Ryan, S. Wheeler, and T. Turner. Impact of Correlations in Logistics Supply Chains. In L. Caccetta and V. Rehbock, editors, *Proceedings ASOR Conference*, pages 146–153, Curtin University of Technology, Perth, Australia, 2005.

[325] S. Wheeler, A. J. Ryan, and T. Turner. Modelling Logistics Chains for Optimal Supply and Maximum Throughput. Technical report, DSTO, 2006.

Even though [24, 139, 209, 271, 325] have not contributed directly to this thesis, my involvement in a wide range of complex systems modelling studies provided an essential grounding for my contributions to the theory of complex systems.

Abstract

The design and management of organised systems, comprised of dynamic interdependent collectives of autonomous agents, is the kind of problem that the discipline of complex systems is intended to address. Nevertheless, conventional model-based applications of complex systems may be of limited utility when the problem is also data-poor and soft. In this case, a quantitative model may be at best meaningless; at worst harmful. Systems approaches, such as soft systems methodologies, have been developed that provide some guidance in this domain. However, these alternatives do not utilise the exact techniques of complex systems, preferring to abandon mathematical representations altogether. It is the aim of this thesis to advance a "conceptual analysis" approach to complex systems design that exploits deep insights from the mathematics of complex systems, without building explicit models of the underlying system. It is argued that this approach can extend the domain of applicability of the discipline of complex systems into situations where quantitative data is unavailable, and human and social factors are significant.

Conceptual analysis of complex systems is inherently multidisciplinary, because it is broader than the foundations of any single conventional discipline. This is reflected in the structure of this thesis, which spans the philosophy, theory and application of complex systems. Part I on systems philosophy develops an understanding of representation, which sheds light on the utility and limitations of models. The history of the systems movement is then surveyed, systemism is distinguished from both individualism and holism, and 'system' is defined. Complex systems is contrasted with both early systems theory and contemporary systems approaches.

Part II on complex systems theory firstly relates the major theoretical concepts within a rigourous information theoretical framework. They include complexity, edge of chaos, self-organisation, emergence, adaptation, evolution and self referentiality. The central systems concept – emergence – is then examined in depth beyond its information theoretic interpretation, leading to a concise definition of emergent properties and emergence. A new framework for understanding emergence in terms of scope, resolution and state yields substantial novel insights. It is shown that emergence is coupled to scope, in contrast to the conventional explanation that relates levels of description.

Part III applies the preceding philosophical and theoretical framework to real-world problems in the defence and security arena. In the first example, the theory of multi-scale complexity reveals structural impediments to success for conventional force structures in asymmetric warfare, such as Operation Iraqi Freedom. The second example analyses the capability development process, which is responsible for transforming the security needs of Government into equipment purchasing decisions. The analysis produces practical recommendations for improvements that address the underlying complexity of the problem. Reflections in the conclusion of this thesis focus on the interrelations between philosophy, theory and application. As the individual contributions of this thesis are woven into a single tapestry, they demonstrate the utility of a multidisciplinary approach to complex systems design.

Foreword

> *Mathematics is not a careful march down a well-cleared highway, but a journey into a strange wilderness, where the explorers often get lost. Rigour should be a signal to the historian that the maps have been made, and the real explorers have gone elsewhere.*
>
> William S. Anglin

Whilst this thesis and my background are notionally in applied mathematics, I think the multidisciplinary label is more appropriate. The interdisciplinary passport of complex systems has provided me with the freedom to cross disciplinary boundaries frequently enough that distinctions between philosophy, mathematics and science seem to be as much a matter of style as substance. One of the great drawbacks of the expansion of human knowledge is the proportional pressure on the individual to specialise in only one "language game", when any messy real world problem requires contributions across many specialties to uncover effective solutions. Even real world problem solving, or "action research", is itself a specialty! The compartmentalisation of knowledge into disjoint pigeon-holes by the Academic System has produced legions of subject matter experts, armed with vocal barrages of well-rehearsed critiques, ready to defend their turf against the occasional generalist and more frequently migrating interdisciplinarian (often a disillusioned physicist) that dares to "swim outside their lane". A systemic cultural bias towards depth over breadth is just one of the barriers to multidisciplinary research.

To me, the great attraction of complex systems is the opportunity to learn in a direction lateral to the usual decomposition of scientific knowledge. This is still a process of specialisation, but the complex systems framework provides significantly more wriggle-space, as well as a fresh perspective on some of the most exciting challenges in science and engineering. I share with Gell-Man the belief that complex systems is capable of constructing bridges between the specialised representations of relatively isolated scientific disciplines. These links have the potential to improve communication and understanding, thereby strengthening the web of science. Complex systems can also help to interpret the relationships between the sciences, by accounting for emergence in a way that is neither reductionist nor holist – a topic which I address in this thesis.

Complex systems is a field that can present itself as an aspiring politician, enigmatic in its attempt to be all things to all people. Popular science books on complexity enjoy a higher profile than seminal papers in the field, and are often a researcher's first exposure and primary impression of complex systems. The success of the likes of Waldrop [319], Holland [153, 154], Lewin [204] and Casti [76] in enthusiastically informing the public about complexity is to be applauded. However, a brief survey of the foundational concepts: complexity, emergence, adaptation, self-organisation and co-evolution, reveals not just a lack of agreement regarding definitions, but often a reliance on definition by intuition or by example. Such definitions either assert that we know Complexity when we see it (except when we disagree), or provide a list of things and proclaim them to be Complex. Neither strategy involves any justification of the definition, and both are vacuous. Any reader who accepts these concepts is really only accepting the preconceptions they brought to the reading[1], which implies that different readers will interpret the same concept in as many ways as there are different backgrounds. This is symptomatic of any field in its infancy, but clarity must be a priority for the research agenda of complex systems. This is not to say that complex systems requires a single foundation or perspective, but to emphasise that the danger of not having well-defined concepts is incoherence. The hyperbole surrounding complex systems buzzwords and catch phrases, none more prevalent and misleading than the "edge of chaos", can only result in the misapplication of complex systems models out of context without understanding the assumptions of the original models. This aspect of complex systems is most manifest within the "complexity management" literature, where a number of gurus peddle management processes using complexity as a metaphor[2].

At the other end of the spectrum, some traditionally hard – and often labelled reductionist – areas of science such as statistical mechanics are associating their research with complex systems. This is puzzling when, according to Bar-Yam [25, p. 91], a complex system can be defined as a system where the assumptions of thermodynamics do not hold. One could rather cynically suggest the involvement of theoretical physicists in complexity is motivated by the funding that fashionable interdisciplinary research attracts, rather than a genuine commitment to emergence and self-organisation, or agent based modelling. However, another way of viewing this development is that statistical mechanics is starting to ask questions beyond the properties of simple ergodic materials in equilibrium, moving from 'thermostatics' towards a genuine theory of thermodynamics. As more complex structures can be treated within a statistical mechanical framework, these tools will provide a new understanding of how systems can increase in organisation and persist far from equilibrium.

Between these two extremes, a remarkably diverse array of research is conducted under the auspices of complex systems, and this diversity is a source of great strength. I like to think of complex systems as almost orthogonal to and entangled with disciplinary science, which suggests that any attempt to demarcate simple versus complex systems is ill conceived. Instead, it is better to view complex systems as specialising in the study of organisation and relationships, in the global implications of structure at multiple scales. Some simple structures are already well treated within disciplinary science, in which case the only contribution of complex

[1] One is reminded of Sartre's [274, p 17] critique of Marx: "I found everything perfectly clear, and I really understood absolutely nothing. To understand is to change, to go beyond oneself. This reading did not change me."

[2] See for example [86, 221, 273]. For more insightful discussions of management theory and complexity see the journal Emergence:Complexity and Organisation.

systems is to make connections with equivalent structures in other disciplines that may benefit from analogous techniques.

But the real work for complex systems is to extend our knowledge of how structures can change in organisation, and how increasingly complex structures can continue to be analysed even when traditional techniques fail. General insights into the nature of organisation, including appropriate mathematical techniques – which must always be supplemented with discipline-specific knowledge – can enable disciplinary science to progress beyond simple models and analytical techniques that fail to fully capture structures in space and time. When someone says that x is a complex system, I interpret this as a claim that conventional approaches to modelling the system have failed to capture an aspect of the system's behaviour that is considered essential to the modeller. Consequently, when at least one significant conventional assumption is revised, this should count as a complex systems approach. And when a complex systems approach is generalised beyond the confines of its original discipline, this should count as a contribution to complex systems theory.

Throughout my thesis I have tried to balance application with inquiry. Application, in the spirit of the engineer, seeks to proceed with building more concrete instantiations of ideas in order to evaluate the ability of new theory to make a discernable difference to people's lives. Inquiry, in the spirit of the philosopher, seeks to question how the framework the engineer builds to help solve problems may constrain thinking and create its own set of problems. Without a deep understanding of both perspectives, the increasingly specialised armchair of contemporary philosophy may appear to the engineer to have nothing to contribute to real-world problems. Meanwhile, science and engineering may appear to the philosopher as merely a new dogmatic techno-utopian cult[3], a secularised mysticism that places Mother Nature in the vacuum left by the exorcism of God from philosophy beginning with Hume. Unfortunately, specialists at either end of the applied/inquiry spectrum commonly hold these misconceptions – perhaps to justify claims for the superiority of their own perspective. Of course, we all know that inquiry and application are opposite sides of the knowledge development coin; it is only because the System encourages specialisation that their relationship is viewed in predominantly competitive terms.

So this is the **Unique Selling Point** of my thesis: I take an end-to-end systems view of the development of knowledge, from philosophy to design via complex systems theory. Thus my approach is inherently multidisciplinary, because it develops knowledge about the whole that is distinct from the knowledge contained in each of the disciplinary parts. Whilst each part of my thesis could have been developed to a greater degree of sophistication by specialising in only that area, this would necessarily sever the connections between the philosophy and application of complex systems that are surfaced in a multidisciplinary treatment. My **Motivation** can be summarised as: to develop novel representations and theories that have an impact on how real world systems are designed and implemented.

On **Method**, where possible I have chosen to align the questions I ask with the "exact sciences", and I expect substantial grounds to be given before directly

[3] I owe this turn of phrase to Darryn Reid.

contradicting working scientific hypotheses. This is a methodological decision: it is not based on metaphysical belief in the entities of the exact sciences, but on the belief that to be anti-scientific – by erecting barriers to scientific investigation – is almost always unhelpful. At the same time, science is not reflexive: it cannot be used to legitimate itself. Additionally, in my chosen application domain – defence science – the most important problems are messy, data poor and inherently soft. In order to discuss foundations and address real world problems, it will be necessary to operate outside the comfort zone of exact science, and be satisfied with qualitative or conceptual analysis. A commitment to multidisciplinarity also demands that the temptation to reduce all explanations to a single foundational perspective is staved off.

On **Convention**, I will depart from the scientific style of the impartial third person: 'we conclude', 'this study opposes', etc., for two reasons. On purely pragmatic grounds, distinguishing between my personal research and conclusions, and the research that 'we' have jointly undertaken – my collaborators and I – clarifies my contribution for the examiner. Also for clarity, instead of using 'we' to refer to 'the readers and/or I' eg. 'we may ask', I will write 'one may ask'. The deeper reason behind this departure, is because the process of knowledge development is inherently personal – it depends essentially on personal judgement. This can be papered over and obscured by impersonal reporting, which only serves to promote a positivist interpretation of science.

In the interests of clear communication I have included precise plain English definitions of technical terms in the Glossary, the utility of which hopefully exceeds its dryness. When a new term is introduced that has a technical meaning, it is <u>underlined</u>, as are cross-references within the Glossary. Use of the Glossary is entirely optional, since the more important definitions are also developed in the body of the thesis.

Contents

List of Tables

List of Figures

Chapter 1

The Model-based Approach to Complex Systems

No self-respecting thesis on complex systems would fail to include an agent based model. This chapter presents a typical agent based simulation of a complex socio-technical system, comparing the effectiveness of direct and indirect control measures in the so-called Global War on Terror. Conclusions are reached regarding the effectiveness of control in the face of naturally arising feedback loops, which are conceptual devices that attempt to explain the symptom of unintended consequences. I then pause to critique this study, which raises the six questions that motivate my thesis.

1.1 Introduction

A journey of two hundred pages begins with a single study. This study is the first agent based model that I published during my candidature, and one that I found increasingly difficult to retain within the body of my thesis, as my subsequent research revealed a host of naïve presuppositions my well-intentioned study had unwittingly incorporated. Even if the more superficial problems could be addressed by rewording the explanatory text, I remained uncomfortable with the model itself. This prelude provides a fitting home for this study. I hope it will become apparent to the reader that my simple model raises many more questions than it answers. These questions provide the motivation for my thesis. I hope that you will find the six questions I promise to address sufficiently compelling to accompany me on my journey through philosophy and complex systems theory, towards an alternative approach to real world applications.

1.1.1 Novel contributions

The study presented in the first section of this chapter is based on:

[269] A. J. Ryan and A. Grisogono. Hybrid Complex Adaptive Engineered Systems: A Case Study in Defence. *Interjournal*, 2004.

Our original contributions are:

1. A brief historical study of unintended consequences in defence and security;
2. An agent based model of terrorism demonstrating that direct control can produce unintended consequences, which may be avoided by implementing indirect controls.

My role in this work was:

1. After the initial meeting to set the scope of the paper, I performed all of this work.

The critique I present in the second half of this chapter has not been published elsewhere. The positive contributions of my critique, which arise from addressing the questions it poses, will constitute the majority of my thesis.

1.2 Unintended Consequences in Defence Operations

> *In the tumult and uproar the battle seems chaotic, but there is no disorder; the troops appear to be milling about in circles but cannot be defeated. An apparent confusion is a product of good order... Order or disorder depends on organization.*
>
> Sun Tzu

While the beginnings of understanding warfare as a complex adaptive systems dates more than 2500 years to the writings of Sun Tzu, recently a growing body of literature describes the broader aspects of defence systems and operations in terms of complex systems science [155, 165–167, 199, 200, 227, 262, 276]. Previous work [139] illustrates that a defence force exhibits the basic properties of a complex adaptive system, and identifies adaptive mechanisms in natural systems at the levels of organism, species and society. Analogous mechanisms exist within defence systems at the levels of adaptive systems, capability development and defence within society, which facilitate adaptation through learning, evolving and cultural change respectively. The defining difference for designed complex adaptive systems is the existence of an externally imposed purpose. Therefore the concept of fitness for designed systems is determined not by their ability to survive and propagate, but to fulfil that purpose - an abstract concept that can itself evolve over time.

The defence systems under consideration are distributed in nature and highly autonomous, yet operate under the centralised control of human commanders and evolve according to the intentions and perceptions of their human designers. This study explores the interactions between centrally imposed controls and the underlying complex adaptive system. Issues of predictability and unintended consequences are discussed with reference to historical studies of warfare. A contemporary terrorist scenario is developed as an Agent Based Model (ABM) to explore the effect of constraining local interaction rules through centralised control. Two alternative control mechanisms are contrasted in terms of their effect on system behaviour.

1.2.1 Design issues

There are a number of issues related to the design of complex systems. Lucas [213] identifies a list of potential problems for implementing complexity theory as a computing paradigm, many of which are relevant to designing complex systems in defence. One problem is how to identify and repair inappropriately evolved systems, meaning those systems that fail to meet their design objectives. This is clearly a valid concern, since brains can suffer mental illness, species become extinct and market economies crash. In each case, the feedback mechanisms have pushed the system into an attractor state (or path/transient) that is adverse to some fitness concept (such as continued survival or propagation) and unless these mechanisms are altered or supplemented there is nothing to indicate that system performance will revert to a more favourable attractor. Most systems in this state require external interventions in order to restore preferred behavioural patterns. For a defence system this role could be performed by the system designer, similar to the way a reserve bank monitors a national economy and adjusts the official interest rate.

Note that this identifies a requirement for continual monitoring and feedback, as well as appropriate mechanisms for system repair when defective behaviour is identified. The way these types of mechanisms could interact with the underlying complex adaptive system is not currently well understood, and controlling a complex system such as an economy or a defence force by altering a single variable is both difficult and unreliable. Central to this problem, and to the most important issues relating to the design and control of complex adaptive engineered systems, are two concepts: unpredictability and unintended consequences.

Unpredictability

Predictability is often considered an imperative for military systems, since warfare can exhibit instabilities characteristic of chaos: microscopic state differences can lead to macroscopic state changes over relatively short time scales. This is the meaning behind the term strategic corporal coined by General Krulak [190]. This describes how the decisions made by a Corporal in a humanitarian aid situation can within a matter of minutes either lead to the defusing of a potentially violent situation, or if the wrong decisions are made, to far-reaching strategic effects capable of jeopardising the mission. An essential role of training, doctrine, rules of engagement and much of military command and control then, is to suppress variation that can potentially produce undesirable macroscopic state change. Taken to an extreme, this urge to suppress variation and constrain the available degrees of freedom can lead to a very brittle and unstable solution: as the strategic environment changes, the lack of variation inhibits the force's ability to adapt. The brittle nature of fixed solutions is a significant contributor to the central role of the 'need for variation' in evolutionary theory.

Given that in the limit, enforcing predictability may come at the expense of robustness, it may be instructive to review the use of inherently unpredictable systems. A simple example is the use of horses in warfare. Since we lack a deep understanding of the cognitive processes of the horse, the behaviour of a 500kg horse that has been spooked is both unpredictable and dangerous. Nevertheless they played a decisive role in combat for over five thousand years up to and including World War I, epitomised by the capture of Atahuallpa by Pizarro in 1532 [109]. Rodney Brooks [65] suggests this is because people have developed an understanding of the parameters under which horses behave favourably. Provided the horse is well fed and we don't light a match next to its eye, a level of trust can be established within this set of parameters through the trial and error of experience, even though we cannot predict the behaviour of an individual horse in a particular situation. Therefore a loss of certainty in system behaviour can be justified provided other benefits, such as robustness or decisiveness in combat, can be realised. Clausewitz [85] argued at great length and with great sophistication on the inherently unpredictable nature of warfare. For example, Clausewitz observes that "[w]ar is the province of chance. In no other sphere of human activity must such a margin be left for this intruder. It increases the uncertainty of every circumstance and deranges the course of events". In this light, attention should be focused on understanding rather than suppressing unpredictability.

Unintended consequences

In any complex adaptive system where centralised control is imposed in order to improve some measure of system performance, it is common to observe a raft of unintended consequences. In some cases the unintended effects of change designed to improve a measure can actually result in a decline. This can be explained in terms of naturally arising feedback loops. The following recorded historical examples will illustrate how formal adaptive mechanisms have been hampered by naturally arising feedback loops.

Artillery in the Battle of the Somme

An account of the Battle of the Somme during World War I from a technological perspective is given in [63]. The firepower of the newly deployed machine gun was so devastating that both sides were forced underground into trenches extending from Switzerland to the North Sea, protected by hundreds of thousands of miles of barbed wire and other obstacles. The tactical problem Allied high command faced was one of penetration to end the stalemate. The solution they pursued for two and a half years as the key to success was prolonged bombardment. A belief in this solution persisted despite repeated failures, such as the Third Battle of Ypres in 1917. A year's production of shells for 55,000 war workers were fired over nineteen days, yet only forty-five square miles were taken in five months at a cost of 370,000 men. To understand why this strategy ended in mass slaughter rather than the desired penetration it is necessary to identify the unintended consequences. Firstly the bombardment shattered the ground, and rain turned the battlefield into a "chaos of mud and swamp" [63], making penetration more difficult, rather than easier. Secondly, the bombardment signalled the intention to advance, meaning surprise was impossible and enemy reserves were ready in place. In combination, these features of the system produced an effect opposite to the intended effect.

Moral hazard and bias in peacekeeping

The issues of moral hazard and bias in peacekeeping and humanitarian aid situations relates to the concept of unintended consequences. The issue of moral hazard is summarised by the question: can actions taken by third parties to forestall violence actually encourage it? Carment [74] suggests that belligerents utilise the security, food and medical aid provided by third parties as a public good. In other words, the services provided by peacekeeping organisations help not only the victims of conflict, but can also be utilised by the perpetrators. Belligerents can access the food and medical supplies, and be protected from retaliation, provided that they are not too easily distinguished from the local population. This reduces the risk of combat and increases the incentive to pursue gains through violence. The bias of taking sides during intervention can encourage violence and result in escalation, such as the NATO involvement in Kosovo. In this case, NATO statements bolstered the Kosovo Liberation Army (KLA) and resulted in military growth, strengthened numbers and increased international legitimacy. The KLA then used ceasefire talks to improve their fighting capacity and reoccupy territory lost in the previous year [74]. Once again, the effect of changing the system is opposite to the intended effect.

The Pentropic Army

In 1956-60 the US Army undertook a restructuring in response to the shift in strategy away from the expense of a large ground force towards a nuclear policy [177]. The Pentomic division was designed to withstand nuclear attack through dispersion, but able to mass quickly to employ a tactical nuclear response. Australia implemented a similar model called the Pentropic division in 1960-65 with the intention of improving interoperability with Australia's main ally, and improving the Army's capability in the nuclear age. However, both interoperability and capability decreased under the restructure. Roger Lee [201], head of the Australian Army History Unit, explains how this occurred:

> Current operations in Malaysia as part of the conventionally structured British Commonwealth Far East Strategic Reserve required battalions to reorganise from Pentropic to standard British Commonwealth battalion structure before service. In addition, the US Army abandoned the Pentomic divisional structure in 1961, leaving Australia interoperable with none of its allies. The restructuring of the Pentropic battalions, now commanded by a full Colonel, resulted in a large rank gap between the commander and the OCs (Officer Commanding) of the supporting elements. Consequently, OCs did not voice their concerns to the Battalion commander, resulting in command failures. Other problems were that a full Colonel may be too old to cope with a field command, and at the next higher command level, the battle group, there was too wide a span of command. Finally, the demise of the CMF Army Reserve, once an important capability, has been attributed to the introduction of the Pentropic division.

An accumulation of unintended consequences overwhelmed any positive effects of the restructure and the Pentropic structure was abandoned in 1965.

1.2.2 Agent based model

The previous section identified different instances of centralised changes to complex adaptive systems producing unpredicted effects on system behaviour, often for the worse. The obvious question is: can centralised control shape emergent behaviour to improve system performance? In this section, an Agent Based Model (ABM) is developed to explore how centralised control influences natural interactions at the individual agent level. The ABM can be considered a distillation, meaning the minimum set of features of a real world situation are included in the model. This is both the greatest strength and weakness of the distillation approach to agent based modelling. An austere representation provides the clearest understanding of the dynamics of the mechanisms under investigation, in the same way as an experimentalist will go to great lengths to construct a laboratory environment where external influences are removed. The disadvantage of this approach is the particular insights cannot be directly applied to the real world without consideration of many complicating subtleties. A large number of equally valid[1] simple models can be derived from the same complex real world situation. A considerable limitation of all ABMs is the implementation of human decision-making as a fixed rule set. Even if heuristic learning algorithms and adaptive rules are incorporated, it is clear that the current state of the art in Artificial Intelligence falls well short of human abilities for cognition, intelligence, invention and adaptation, all of which affect the behaviour of socio-technical systems.

Terrorism scenario

We will use a hypothetical anti-terrorism scenario to investigate unintended consequences in complex systems. An issue-motivated group is using violence as a method for achieving its goals. The indiscriminate use of violence against civilian targets has led to the group being referred to as terrorists in the international media. The government is under pressure to resolve the situation quickly with the minimum bloodshed, and the terrorists are not open to negotiation. Terrorist bombings have led to public outrage and calls for retaliation. However, retaliatory strikes seem to only have increased the terrorists' resolve and led to a cycle of violence with civilians caught in the crossfire. Military options that provide the required level of survivability against armed terrorist threats cannot guarantee a level of precision capable of preventing innocent casualties.

An alternative approach has been tabled in parliament. An immediate cessation of retaliatory strikes combined with a doubling of police presence will discourage the gathering of terrorists. It is argued that over time this will reverse the negative publicity of civilian casualties, and as the population rejects the violent methods of the terrorists, their support and recruiting base will contract.

Terrorism model

A model is now developed in order to contrast the control mechanisms representing the two approaches to the terrorist problem outlined in the previous section. The terrorist scenario is implemented in NetLogo [328] as a multi agent system. Interaction rules at the agent level are specified and aggregate system behaviour is described as being emergent from the local interactions. As the parameter space is explored by changing interactions it is possible to measure a change in the emergent behaviour of the system state over time. The local interaction rules are described in Table 1.1 below. The model contains five types of entities: civilians, terrorists, police (these could represent a police or military peacekeeping force), aerial weapon systems (these could equally represent fighter jets, homing missiles or attack helicopters, and are referred to as jets) and bombs. Civilians and terrorists live for 70 turns, and police numbers are kept constant. All new entities are born as civilians and terrorists must recruit from the civilian population to increase in number. The civilian birth rate is approximately equal to the death rate. Movement for all entities except jets is linear in a random direction. The

[1] Here, 'equally valid' is meant in the sense that no one simple model is intrinsically more privileged than others.

jets travel at three times the speed of other units and track a particular target. When a jet drops a bomb it explodes immediately. Terrorist bombs explode each turn with a 20% chance. This means the probability a bomb kills the terrorist who planted it is the probability the terrorist does not escape the blast radius of three squares (terrain is represented by square patches arranged in a two dimensional lattice). This equals $1 - (1 - 0.2)^3 = 0.488$, which means that 49% of terrorists are suicide bombers. In the case that the terrorist escapes the explosion, a jet is dispatched to bomb the terrorist.

Table 1.1: Local interaction rules for a NetLogo agent based model.

Interacting Entities		**Interaction**	**Rule**
Terrorist	Bomb	Plant	Plants a bomb every 10 turns
Terrorist	Civilian	Recruit	10% chance of recruiting neighbouring civilians
Police	Terrorist	Discourage	Terrorist will not recruit civilians if police are present
Bomb	All except jets	Kill	All units in radius 3 die
Jet	Terrorist	Bomb	If a terrorist bomb explodes, the responsible terrorist is tracked and bombed
Civilian	Jet Bomb	Convert	If a jet bomb causes civilian casualties, all witnesses are converted into terrorists

The model captures each feature of the scenario as a number of simple local interaction rules. Figure 1.1 below illustrates a NetLogo simulation in progress. Jets and police are coloured blue, civilians are coloured white, bombs are coloured black and terrorists are coloured red. Terrorist bomb explosions are red, while explosions of bombs dropped by jets are pink. If jet bomb explosions cause any civilian casualties, the civilians who witness the explosion (within a radius of two squares) are converted into terrorists and displayed in pink, rather than red. The initial settings are 100 civilians, 10 police and 10 terrorists. A jet entity is created each time there is a new mission. This assumes that there are more jets available to the government than retaliation missions at all times.

To establish the effect of different local interaction rules on the level of terrorist activity, the model is replicated twenty times for four parameter settings. The first setting is the current policy of retaliation against terrorist bombers and the current police levels. The second setting doubles the number of police while continuing retaliation. The third setting does not launch retaliation strikes against terrorist bombers and maintains current police levels. The fourth setting both removes retaliation and doubles police staffing. A simulation is run for each setting for 1000 time steps or until the number of terrorists reaches zero. The system variables measured to gauge the success of different settings are the number of bomb explosions and the number of lives lost from both terrorist and jet activity after 1000 steps, the average number of terrorists alive per time step, the average number of

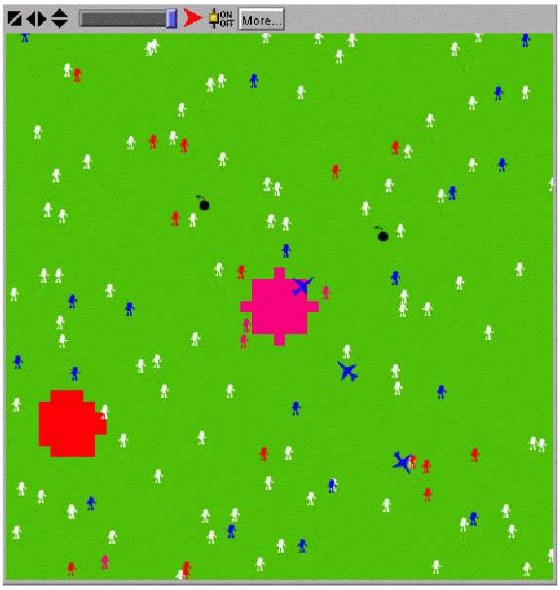

Figure 1.1: NetLogo agent based model of a terrorism scenario.

time steps each replication runs for, and the number of runs where there are no terrorists left after 1000 time steps. The results averaged over twenty replications are shown in Table 1.2 below.

Table 1.2: Results of 20 NetLogo replications.

	Setting 1	**Setting 2**	**Setting 3**	**Setting 4**
	Retaliation Police = 10	Retaliation Police = 20	No Retal. Police = 10	No Retal. Police = 20
Terrorist bombs	711	382	172	90
Terrorist kills	2307	1243	530	271
Jet bombs	254	141	0	0
Jet kills	743	412	0	0
Num. terrorists	30	19	11	8
Av. time Steps	930	678	431	308
Num. runs where terrorists die out	3	11	16	20

When the model is replicated for setting 2 and compared to setting 1 (see Table 1.2 above), both the number of bombs planted and number of casualties from bomb explosions are reduced by over 45% by doubling the police presence. The average number of terrorists per time step is reduced from 30 to 19, although for both settings this is still greater than the simulation's starting value of 10 terrorists, which shows a growth in the number of terrorists over time. The number of retaliatory strikes reduced from 254 to 141, resulting in 218 less casualties. For Setting 1 the terrorists died out in only 3 of the 20 replications within 1000 iterations, and the average replication length was 930 time steps. With police numbers doubled to 20 in Setting 2, 11 replications terminated and the average replication length was reduced to 678 steps. From this data we can conclude that the police presence has had a significant positive impact on the level of terrorist activity, the number of retaliatory strikes required and the likelihood that terrorism is eliminated completely. This result could be expected, although it is interesting that inhibiting one mechanism for recruiting new terrorists has such a large impact.

Less intuitive is that by prohibiting targeted bombing of terrorists through retaliatory strikes, the average number of terrorists, and therefore the average time steps, the number of terrorist bombings and killings all reduce. For Setting 3, the improvements are greater than Setting 2. The reason is that although the jet bombings target and kill on average 254 terrorists, as a side effect it converts an average of 299 civilians into terrorists. These extra 45 terrorists over the life of the simulation represent 23% of the terrorist population, which has a multiplier effect to increase terrorist killings by 77% from 530 to 2307. In Setting 4, which prohibits retaliation and doubles the police presence, the best results are achieved. The number of bombings is reduced to 90, the average number of terrorists is 8, which is lower than the initial conditions, and all 20 replications terminate within 1000 time steps. Statistically, using a two-sample single tailed *t-procedure*, the difference in terrorist activity (measured by the number of exploded terrorist bombs) between Setting 1 and Setting 4 is significant at the 99.95% confidence level, and the t-statistic is 9.82, which is much larger than the critical value of 3.88. Despite the high significance, a note of caution should be raised at this result, since the data is not normally distributed and contains outliers, which affects the validity of the *t-procedure* for a sample size of twenty replications. Other analyses of ABMs [199, 200] suggest the data displays fractal characteristics due to correlations between entities in both space and time. Consequently, modelling the ensemble as a statistically scaling system may be more appropriate.

Conclusions

The ABM simulates two mechanisms for centralised control within a complex adaptive system. In the model, autonomous agents make decisions based on local interactions. Both mechanisms are intended to reduce the number of terrorists. The retaliation mechanism, which operates by directly reducing the number of terrorists, is subverted by local interactions. Consequently, it has the opposite effect of increasing the number of terrorists and bombings. The policing mechanism is less direct, since it does not kill or detain existing terrorists, nor prevent them from bombing. However, it does inhibit the recruitment of new terrorists, leading to decay in the terrorist population over time. The policing mechanism can be understood to be operating on a longer time scale over several generations of terrorists, rather than the short-term strategy of retaliating against individual terrorists. The model is intentionally simple compared to the historical examples discussed earlier and only contains a small number of feedback loops. As more interaction types are added to the model, interference between feedback loops may increase the complexity of the model's behaviour. Many more counter-intuitive effects of centralised control measures may be revealed, but this model demonstrates a low complexity threshold for the emergence of unintended consequences.

Most importantly, this model illustrates how indirect action altering the context in which local interactions take place can be more effective than centrally imposed policies focussed on directly improving a single system measure or benchmark in complex systems.

1.2.3 Summary

In designed complex systems, human design and engineering mechanisms interact with the naturally arising adaptive mechanisms, such that only limited control is possible. Exploring the implications of conscious human design decisions, the possibility for spontaneously arising informal adaptive feedback loops to undermine deliberate, formal adaptive mechanisms was demonstrated. The potential for designed systems to harness the benefits of the underlying adaptive systems while utilising intelligent design effort to shape the emergent system behaviour was illustrated using an agent based model. The results may provide a basis for a new understanding for effective change within complex systems, with implications for system design, capability development and force transformation within defence.

1.3 Critique, or Taking the Red Pill

> *This is your last chance. After this, there is no turning back. You take the blue pill – the story ends, you wake up in bed and believe whatever you want to believe. You take the red pill – you stay in Wonderland and I show you how deep the rabbit-hole goes.*
>
> Morpheus in The Matrix

The model I have presented was well received at the International Conference on Complex Systems in 2004[2]. It was also influential on the direction of my candidature, because it led to the clear decision that I did not want to simply turn the handle of an ABM-making machine to produce a parade of complex systems models. I sensed that there were deeper and far more important questions beneath the facade of my simple model and its statistically justified conclusions. This section critiques my study of unintended consequences, revealing a number of assumptions that deserve further examination.

The opening sentence of the study immediately raises a question: what does it mean to understand warfare "as a complex adaptive system"? Does it imply that Complex Adaptive System (CAS) is a category, or a type of thing? Could we classify parts of the world as being a CAS? If so, what other categories are there? Is CAS the subset of complex systems that have the capacity to adapt, and in turn are complex systems the subset of systems that are complex? What things are not systems? All of these questions make the same mistake of confusing the way an observer structures their thinking about the world, with the world itself.

[2] The only constructive criticism I received was that the model could have been more realistic – that it would be better if I had the 'real' feedback loops in the model. The critique I present here will go well beyond questions of model accuracy.

This mistake is most easily revealed when one realises that it is the observer that chooses the boundary of the system. As I will discuss in detail in Chapter 3, a system is not an objective region of space-time, it is an approach for structuring our understanding of the world.

The first two paragraphs also introduce a bundle of closely related concepts: fitness, purpose, intention, perception, autonomy, predictability and control. Again there is no distinction between concepts that observers use to explain the system, compared with intrinsic features of the dynamics of a process. These concepts will require some work to untangle, but for now I will point out that talk of "externally imposed purpose" implies that the designer is outside the system boundary, yet to "evolve according to the intentions and perceptions of their human designers" is to bring the designer back inside the system boundary. Two things are clear: my study has not been clear or consistent regarding boundaries or terminology; and if the internal models and perceptions of agents affect their behaviour, this needs to be accounted for when developing an understanding of systems.

The section on design issues provides an explanation for systems that have evolved inappropriately[3]. It claims that "the feedback mechanisms have pushed the system into an attractor state that is adverse to some fitness concept". This claim is extremely general, supposedly accounting for the behaviour of brains, ecosystems and economies. But what does it mean for a feedback mechanism to push? What are the entities involved in the feedback relation? In most system dynamics models of feedback, these entities are abstractions and aggregations, not spatially located physical objects. Fitness is not the only conceptual device in this explanation: attractors and feedback mechanisms are also just concepts for explaining observed behaviour. They are part of the systems perspective. But this begs the questions of what constitutes a systems approach; and what are the relationships between the concepts that populate a systems framework.

The historical arguments add some depth to the discussion, because they use a complementary perspective to the systems approach. But on closer examination, even though they provide anecdotal support for the existence of unintended consequences, they give no support to the conjecture that the cause of unintended consequences is feedback loops within a complex adaptive system. Surely there is a better way to integrate the insights of a complex systems approach with existing approaches.

One of the more interesting statements within the terrorism model is that "aggregate system behaviour is described as being emergent from the local interactions". Because behaviour is "described as being emergent", here the distinction is made that emergence is a concept in the eye of the observer. The obvious question this raises is whether emergence can occur independently of an observer. The emergence of life from non-living matter is often taken to be a prototypical case of emergence, and yet surely it is not contingent on any observer. A secondary question is whether any and every combination of local interactions generate emergent behaviour. When is it appropriate to describe aggregate behaviour as emergent? Do the null interaction rules generate emergent behaviour? Is a Gaussian dis-

[3]The idea that something can evolve inappropriately is itself highly questionable, because it confuses a tautologically defined process with the value judgements of an external observer.

tribution an emergent property? If so, how can independent components share collective emergent properties? Alternatively, is emergence only when the aggregate behaviour cannot be described by an equation or model that is simpler than the underlying dynamics? This sort of answer has been suggested in the literature, but it is deeply unsatisfactory, because it depends vitally on the degree of knowledge of the observer. It leads to an association between emergence and surprise, and to an inescapably relativist definition of emergence.

As was foreshadowed in the introduction, perhaps the most problematic aspect of this study is the model itself. Granted, two important limitations are stated up front: the model is overly simple and does not and cannot account for human decision-making. But what impact do these limitations have on the conclusions drawn from the model?

As was acknowledged above, a large number of equally valid simple models can be derived from the same complex real world situation. Consequently, a single model tells us very little about the problem domain. But it gets worse than this. Because agent based models are in general massively over-parameterised, traditional approaches to validation are problematic. This is because given almost any real world data set, one can expect to find at least one set of tuned parameters that reproduces the phenomenology. Therefore, reproducing the observable behaviour of the real world does not imply that the underlying model is valid or generalisable. There are similar issues in analysing the sensitivity of the simulation model to changes in resolution. Whereas the analytical solution to a set of differential equations can provide deep insight on sensitivity that is independent of resolution, the same cannot be said for the result of digital computations. Without running large numbers of models with extremely high precision a statistically significant number of times, the impact of the choice of time step and the precision of each of the model parameters is unknown in agent based models. Chaos theory shows that arbitrarily fine scale differences can be amplified into arbitrarily large deviations over a number of time steps, so the finite granularity of agent based models may be significant – at the very least there is no way to prove that it is not.

Due to the nature of agent based modelling, the simulation requires a precise number (or at least a probability distribution) for every parameter. This often results in quantifying the unquantifiable. For example, in my study a witness who is within two squares of a retaliatory strike killing innocent civilians, always converts to terrorism. The basis for this rule is that the incident combines two ingredients that are expected to be common amongst terrorists: the experience of trauma, and disagreement with the way power is executed within the current political and social structure. However, even if a psychologist agreed that these are necessary conditions for conversion to terrorism, they would hardly be prepared to commit to a quantitative rule that related these causes to effects. The retaliatory strike explaining why one witness joined a terrorist organisation, may equally explain why the next witness joined a peace activist organisation. To say that witnessing the death of innocent civilians is a cause of terrorism is like saying that being born is a cause of terrorism – even if necessarily true it is not revealing. Uniform and fixed simple rules will never capture human decision-making: even when we understand how a society of rational rule-following robots behave, there is much more to know about human behaviour, and this is why social sciences exist. Rule based, mechanistic explanations of human behaviour are manifestly inadequate, given that arguably a defining characteristic of human intelligence is choosing when to break the rules.

In conclusion, the conceptual framework of this study is confused and the terms are not well defined. For the problem domain I seek to address, agent based modelling appears to offer limited insights as a direct and explicit model of the socio-technical system. Some of the issues I have raised in this critique apply specifically to my agent based model, but most hold for any agent based model. Similar problems afflict dynamical systems, system dynamics, cellular automata, social networks, artificial neural networks and other model-based approaches when they are used to explain many types of social phenomena. And yet, if the conceptual framework could be more rigorously developed, there ought to be a way to contribute valuable insights towards understanding and intervening in complex socio-technical systems. This idea is not new: Bar-Yam's [34] recent book applying complex systems to real-world problems takes precisely this approach. However, little attention has been paid so far to the legitimation of a "conceptual analysis" approach towards complex systems design and management.

1.4 Organisation of Thesis

This thesis is organised to address the criticisms from the previous section, so that more appropriate applications of complex systems can be realised. The goal is to present applications based on a sound philosophical foundation and a rigourous theoretical framework.

The critique of the previous section can be reformulated as raising the following six questions:

1. How do models represent the underlying system of interest?
2. What is a systems approach?
3. What sort of insights are unique to a complex systems approach?
4. How are the concepts of complex systems interrelated?
5. What is emergence?
6. Are there alternatives to model-based applications of a complex systems approach?

These are big questions, but as the Australian complex systems researcher David Green is fond of saying, "big questions get big answers". The first three questions are predominantly questions of philosophy. Question one requires an understanding of the utility and limitations of representations, and how they can be useful models of the system of interest. Philosophy of mind and cognitive science provide an extensive literature on representation that is not widely known within complex systems. Question two asks about the foundations of systems thinking, and the philosophy of systems. This is something that analytic philosophy is largely silent on, but that early systems theorists were keen to address, by contrasting their approach with disciplinary science. Question three requires a contrast between complex systems and other systems approaches, in addition to the contrast between systems thinking and science.

Questions four and five are questions of complex systems theory. Often, the distinction between systems that are self-organising, adaptive, and complex is not clear. Nevertheless, this is the kind of question that should have theoretical answers within the discipline of complex systems. Question five concerns *the* central feature of any systems approach, the rather mysterious concept of emergence. Yet there is no theory of emergence, or perhaps more accurately there are many theories of emergence, none of which are satisfactory.

The sixth question is about broadening the application of complex systems. Agent based modelling is paradigmatic of complex systems research, and will continue to play an important role in understanding the dynamics of complex systems. Simple models can generate deep insight: Kauffman's random boolean networks [175], Arthur's El Farol bar [17], and Conway's game of life [130] are notable examples. However, as the previous section shows, there are significant limitations to model-based approaches, even when the insights of complex systems may be relevant to the problem. This suggests that there is room for complementary techniques for the application of complex systems. This is especially true for soft problems, where human factors are significant and quantitative data does not exist.

My thesis is organised in three parts that propose initial answers to these questions. Part I discusses the philosophy underlying complex systems, which is so scarcely discussed that one could be forgiven for presuming it is taboo in the complex systems literature. The consequence of ignoring philosophy is ill defined terms and bold conclusions that are reliant on a trivial relationship between models and the aspects of the real world they are supposed to represent. It is the mistake that Broad [62] masterly summarised as confusing the Author of Nature with the Editor of *Nature*. Part I surveys key philosophical insights – in the broadest sense of the term – that I use to frame my understanding of complex systems. Chapter 2 explores the topic of representation, presenting a theory of the representational relation between entities, models and agents. The development of clear and explicit definitions provides a precise foundation for subsequent chapters. Chapter 3 digs a bit deeper into exactly what is meant by 'a systems approach'. The history of the systems movement is outlined, and a novel definition of 'system' is developed. Chapter 4 presents the case for an alternative to model-based applications of a complex systems approach. This chapter gives broad guidance for conceptual analysis and identifies the most important considerations within this approach.

Part II develops the theory of complex systems. I begin by casting the central concepts of complex systems into information theory in Chapter 5. This facilitates precise definitions and enables an exploration of the relationships between concepts. The concepts I address are:

- Complexity;
- Edge of chaos;
- Self-organisation;

- Adaptation;
- Evolution; and
- Self referentiality.

I then move beyond the information theoretic interpretation for the important concept of emergence. A new framework in terms of scope, resolution and state is developed. In Chapter 6, emergent properties and emergence are formalised within the new framework to provide precise, testable definitions. It is shown that emergent properties are coupled to scope, and there exist fundamental limits on our ability to model emergence in formal systems. Chapter 6 is the major individual contribution of my thesis, because emergence is such a central concept in systems research, and yet to date it has proved so difficult to understand. So many contradictory statements have been made about emergence because it is viewed as both paradoxical and imprecise. I contend that this is inevitable when emergence is explained in terms of hierarchical levels. In light of the new framework, emergent properties are no longer mysterious – they are simply a difference in local and global structure.

Part III applies conceptual analysis to two problems in the defence and security domain. The first example in Chapter 7 applies the theory of multi-scale complexity to the operational imperative of countering asymmetric warfare. Deep structural impediments are revealed, and possible solutions are discussed. The second example in Chapter 8 critiques the capability development process, which is responsible for transforming the security needs of Government into equipment purchasing decisions. This is the most challenging application of complex systems I consider, due to the scale, complexity and socio-technical nature of the problem that the process is designed to address. While quantitative methods are not appropriate, conceptual use of complex systems engineering theory provides a number of significant insights, which are suggestive of practical improvements to the capability development process. Together, these applications demonstrate an alternative to model-based analysis of complex systems, making use of both theory and philosophy to guide practice. The applications are where the value of the preceding inquiry is realised: where deeper understanding can influence the way forces are organised and capability is designed.

The final chapter summarises the results of my thesis, and draws concluding remarks that link together the insights that span my philosophical, theoretical and applied research. It addresses a final question, which is perhaps the most important question: why do multidisciplinary research? I argue that I have demonstrated how the synthesis of philosophy, theory and application can transcend the artificial boundaries that have been erected to insulate the specialist. Moving beyond traditional competitive relationships has opened up many cooperative synergies across disciplines, which is a rich and largely untapped source for novel contributions. Indeed, I believe that the principle novel contribution of this thesis is to show that the novelty within each chapter is a direct result of being conducted within a multidisciplinary context. The concluding chapter reviews the interconnections between the chapters to add weight to the case for multidisciplinary research. In short, my thesis demonstrates the need for multidisciplinary research, because there exists knowledge that lies beyond the reach of individual disciplines.

Part I

Philosophy

Chapter 2

Representation

Representation is inherent to the concept of an agent, but its importance in complex systems has not yet been widely recognised. In this chapter I introduce Peirce's theory of signs, which facilitates a definition of representation in general. In summary, representation means that for some agent, a model is used to stand in for another entity in a way that shapes the behaviour of the agent with respect to that entity. Representation in general is then related to the theories of representation that have developed within different disciplines. I compare theories of representation from metaphysics, military theory and systems theory. Additional complications arise in explaining the special case of mental representations, which is the focus of cognitive science. I consider the dominant theory of cognition – that the brain is a representational device – as well as the sceptical anti-representational response. Finally, I argue that representation distinguishes agents from non-representational objects: agents are objects capable of representation.

2.1 Introduction

Representation is an essential concept for understanding the behaviour of agents in a complex system. Consider traders in a stock exchange market as agents. If every agent has unmediated access to the value of a company (including its exact future profits discounted to present value), then the market cannot exist, since shareholders would only be willing to sell above this value, a price no rational buyer would pay[1]. Only when partial information on value is allowed and different agents have access to different information is it possible to predict the formation of a market. In this case, each agent must construct a model representing the perceived value of a company. By communicating, agents can modify their models to take into account the representations other agents in their social network have constructed. Because there is a benefit in being connected to agents who are better at predicting future value, some agents may specialise in developing predictive models and charging other agents for access to their expectations (such as financial advisors). Markets would not exist if there were not differences between agents in their representations. Variety in representation allows the simultaneous existence of buyers and sellers, as well as the potential for a secondary market based on constructing representations and selling advice. Even though this is quite obvious, imperfect information, bounds on rationality, and consequently the need for constructing representations did not feature in the theories and models of classical economics.

It turns out that an account of representation is just as important in understanding the role of the discipline of complex systems, as for understanding the behaviour of agents within a complex system. This is because, as will be shown in Chapter 3, the systems approach is one way of representing the world. When this is overlooked, systems applications may be blind to the limitations of the representations they employ. The treatment of representation in this chapter is intended to be interpreted on two levels. On one level, when an analyst uses a complex systems approach, they invariably construct systems representations. On another level, when the system contains agents that also represent their environment, this must be accounted for in any model of the system.

Section 2.2 makes the metaphysical assumptions of this chapter explicit. Then in Section 2.3, Peirce's theory of signs is used as a basis for a theory of representation in general. When agents represent their environment, they may use either external or internal models. Section 2.4 surveys accounts of external representation across several disciplines, while Section 2.5 surveys internal representation, which has been discussed mostly in philosophy of mind and cognitive science. This chapter concludes by defining 'agent' in Section 2.6, which demonstrates the strong link between agency and representation.

[1]One might expect trades to be made exactly at the value of the stock. However, once a financial or time cost is included no rational buyer can exist. Why would an agent buy shares that never increase in real value and incur an exit fee?

2.1.1 Novel contributions

I am the sole author of this chapter, which has not been published elsewhere. My original contributions are:

1. A general definition of representation;
2. Development of three categories of external models in representation;
3. A comparison of theories of representation spanning philosophy, military theory and systems theory;
4. Analysis of the representationist/anti-representationist debate, which identifies that while the discourse has centred on representation-as-mirroring, a more appropriate framing is achieved by representation-as-inference; and
5. A definition of agent.

2.2 Metaphysical Assumptions

Before representation is discussed in detail, it is prudent to make the metaphysical assumptions of this chapter explicit. The metaphysical position I will adhere to is known as physicalism, the view that there are no kinds of things other than physical things. In particular, I assume that the relationship between macroscopic and microscopic phenomena is one of supervenience. The Stanford Encyclopedia of Philosophy offers the following definition:
Definition 2.1 (Supervenience). *A set of properties A supervenes upon another set B just in case no two things can differ with respect to A-properties without also differing with respect to their B-properties [220].*

Supervenience, along with physicalism, entails that in principle, all of the bookkeeping regarding forces can be accounted for in purely physical terms between arbitrarily small entities, when the set B is taken to be the properties of fundamental physics. This is because every time there is a change in a macro level property, there must be a corresponding change in the micro level properties. That the physcial forces fully account for the dynamics at the micro level tells us little about what physical predictions *mean*. Semantics is always relative to an agent's subjective experience of the world, a concept which does not feature in, and cannot be fully explained by, the elementary particles and fundamental forces of physics. First-person experience is just one example of an emergent property (see Chapter 6), the general reason why descriptions at other levels cannot be eliminated. I will assume that forces in chemistry, biology, psychology and sociology do not add anything to the physical: that the laws of physics are *conservative*. This is consistent with Anderson's [11] twin assertions that all ordinary matter obeys simple electrodynamics and quantum theory, but that "the ability to reduce everything to simple fundamental laws does not imply that ability to start from those laws and reconstruct the universe". This assumption can be argued with, but it cannot be proved either way. I assume supervenience regarding the relationship between macro and micro phenomena because to do otherwise is to place some entities outside the domain of scientific explanation, and it is difficult to see what is achieved by doing so. Descartes' [107] non-physical mind that provided the basis for substance dualism in Meditations VI, and Bergson's [43] *elan vital* that animated the evolution and development of organisms, are examples of non-physical entities that have been postulated in science, and history suggests both acted as barriers to progress. Consequently, I only consider representations that supervene on the physical as meaningful.

Secondly, the multidisciplinary theme will already be apparent in this chapter by the way that connections are made between conversations occurring in relative isolation in different disciplines. Continual movement between disciplinary frameworks may initially be disconcerting, because each discipline has idiosyncratic terminology that must be introduced in order to retain the precise meaning of the theories expressed within that discipline. Also, multidisciplinarity does not adhere to the conventions of any one discipline, and to be trained in disciplinary research is to become proficient in the application of negative feedback whenever these (often tacit) conventions are breached.

The alternative – to survey the literature of multiple disciplines through the lens of only one set of disciplinary conventions – is what I will refer to as an interdisciplinary[2] approach. While interdisciplinarity has a lower overhead, it only marginally extends beyond the limits of disciplinary knowledge. An assumption underlying my approach is that multidisciplinarity is not subject to the same limits, because it attempts to do more than just apply the one hammer to a broader domain of nails. *Multidisciplinarity cannot rest on the foundation of a single discipline.* The lack of a universal approach for multidisciplinary discourse is discussed extensively by Kline [185]. In order to provide significant insights beyond those that are already well documented within disciplines, I will focus on parallels between disciplinary discourse – not just to identify similarities and differences, but to comprehend features within a broader scope. The associated terminological baggage is an unfortunate but necessary cost of this approach.

2.3 Representation in General

> *Things don't mean: we construct meaning using representational systems – concepts and signs.*
>
> Stuart Hall

There are a number of reasons why unmediated interaction with the world can be undesirable. Some entities are distinctly unfriendly, others are inaccessible, and sometimes the process of interaction is too costly or time consuming. In order to understand anything about a solar flare on the surface of the Sun, mediated access, via the construction of models, is necessary to avoid the undesirable consequences of unmediated contact. A model acts as a representation because it *stands in* for unmediated interaction with the system of interest. Other situations where representations stand in for unmediated interaction include predicting properties of

[2]Unfortunately, the literature is divided as to which of interdisciplinarity and multidisciplinarity is the broader and more novel programme. See for example [185] compared with [8].

previously unrealised configurations; designing artifacts that do not exist; facilitating comparison of structural similarities between apparently dissimilar phenomena; and generalising knowledge to apply beyond a single entity at a single moment in time[3]. As will be discussed in Section 2.5.1, the dominant theory of human cognition assumes that the mind is a representational device, and that the brain has representational content.

In counterpoint to the important and varied roles of representation, there exists little formal work on representation in general. What are the necessary and sufficient conditions for representation to occur? What kinds of representations exist? One such general theory was proposed by Peirce, which he named the theory of signs or "semiotics". However, because much of his work was misplaced and posthumously edited non-chronologically into highly fragmented volumes, and since Peirce's unique and subtle philosophy requires explication before the finer points of his theory of signs can be appreciated, it remains under-utilised as a general theory of representation. Fortunately, Von Eckardt [312, p. 143-159] has performed considerable work to situate Peirce's theory of signs as a foundation for the more specialised debate on mental representation in cognitive science. I will draw heavily on Von Eckardt's interpretation of Peirce, since it is better oriented towards contemporary concerns in the theory of representation. Unlike Peirce or Von Eckardt, my interests apply to the field of complex systems, and so I will abuse the semiotic and cognitive science terminology by translating it into more general language.

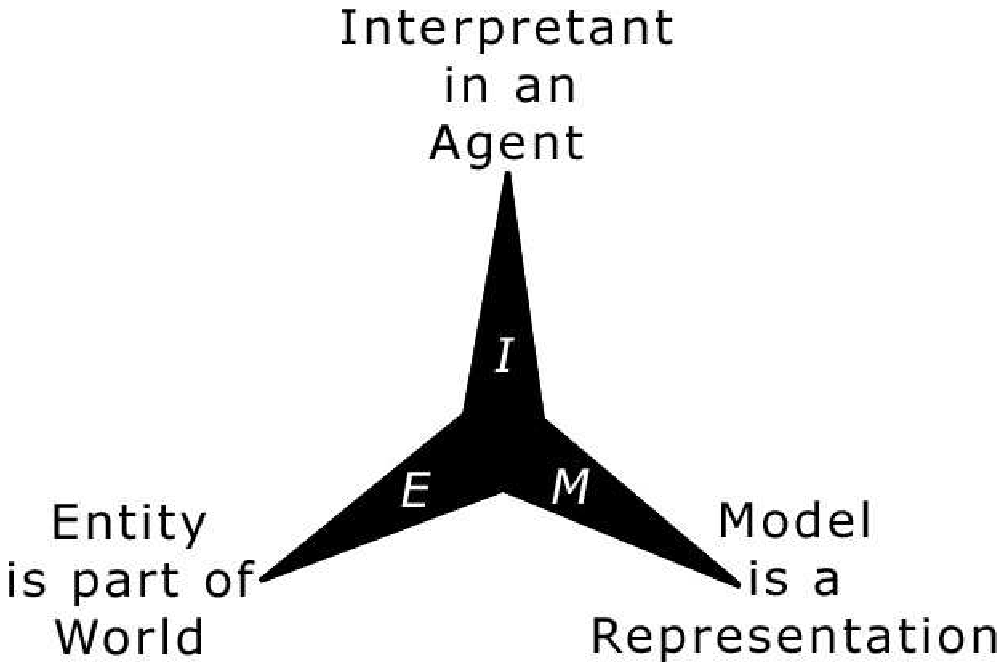

Figure 2.1: Representation as a triadic relation between entity, model and interpretant.

For Peirce, representation was an irreducible triadic relation between objects, signs and interpretants. This implies that something is a sign only if it is a sign *of* an object *with respect to* an interpretant [312, p. 145]. It also implies that the rep-

[3]This list paraphrases Kline [185, p. 19].

resentation relation cannot be decomposed into diadic relations between entities, objects and signs. According to Peirce, representation can only be fully understood by considering the three components of the triadic relation simultaneously.

In the triadic relation, the sign is a token that signifies the object. The more general term I will substitute for sign is model. Peirce's object is already quite general: it may be abstract or concrete, a singular object or a set of objects (a complex object). However, since objects do not in fact need to be objective (or concrete), I will use the more general term entity. The interpretant exists in the mind of the interpreter for whom the sign is a sign [312, p. 148]. Whereas Peirce and Von Eckardt limit their attention to human interpreters, I will generalise this to consider agents. The cost of this generalisation is that the precise nature of interpretants cannot be specified in a way that applies to all agents. Consequently, I can only induce the existence of an interpretant by an observable effect on the agent's behaviour. The triadic relation between entity E, model M and interpretant I is illustrated in Figure 2.1.

Von Eckardt [312, p. 158] summarises the value of Peirce's theory of signs for understanding representation in general as follows:

1. Peirce's distinction between a representation and a representation bearer;
2. His insistence that something can be a full-blown representation only if it is both grounded and interpreted;
3. His attempt to understand what makes a mental effect an interpretant of some particular representation;
4. His struggle with the problem of interpretation for mental representation;
5. The idea that model and entity are related by two very different sets of relations—semantic relations (such as representing, signifying, referring to, and expressing) and the ground relations in virtue of which those semantic relations hold;
6. His taxonomy of kinds of ground; and
7. His apparent interest in ultimately understanding representation in a completely naturalistic way.

Examining each of these points in turn, the first distinction leads Peirce to consider the character of a model itself. Von Eckardt uses the term "representation bearer" to refer to the properties of the model that belong to the model itself, and not to the entity it represents. In a similar vein, Kline [185] brings attention to the essential difference between a model and an entity by invoking what he calls Korzybski's Dictum, after Alfred Korzybski's [188] warning that "the map is not the territory". It is important to remember that Korzybski's Dictum applies to all representations. It implies that a representation must behave differently to the entity in some contexts. Representations are not perfect substitutes, which means there always exist limits to their ability to stand in for the entity they represent. The other implication of this distinction is that it is both possible and useful to understand the properties of representation bearers as distinct from the properties of the entity they represent.

The second item addresses the interpretation and grounds of a model. The reason that the interpretant is a necessary component of the triadic relation is because a representation is more than just a logical similarity between two entities. If a tree casts a shadow, it is not telling the time until an agent *uses it* to tell the time [1]. Or in the words of Dennett, "Nothing is intrinsically a representation of anything; something is a representation only *for* or *to* someone" [104, p. 101]. By shaping the agent's behaviour, I brings M into the appropriate relation as a representation of E. That is, when an agent interprets (and therefore understands) the model, this grounds the model as a representation of E.

The third item links interpretation with a change in behaviour. If I is in agent A, then the representation must be capable of shaping the behaviour of the agent through the presence of I. Peirce classified interpretants as emotional (feelings), energetic (efforts) and logical (habit-changing) effects. According to Von Eckardt, logical effects, which modify the interpreter's disposition to behave, were considered the primary effect.

The fourth item refers to the problem of infinite regress. While non-mental representation is relatively straightforward, issues arise when M is internal to the agent. The difficult question one faces is "what interprets a mental representation?" If mental representations are interpreted in the same way as non-mental representations, this gives rise to an infinite regress of thoughts interpreting thoughts [312, p. 282].

Separating semantic and ground relations, as noted in point five, allows one to account for how a semantic relation can come to exist. In order for a model M to produce an interpretant I in an agent A, it is necessary for A to understand the representation, which requires A to have knowledge of what the ground is. The following example clarifies this account [312, p. 156]:

> For example, suppose I see a photograph. To understand that photograph I must know (in some sense) that there are both a causal relation and a similarity relation between the photograph and its subject, and I must know (in some sense) the respects in which the photograph is a causal effect of and is similar to its subject. If I know all that, then I will be able to form a belief or a thought about the subject of the photograph (that is, who or what the photograph represents)—specifically, that there was such a subject and that this subject looked a certain way at the time the photograph was taken. In other words, by considering the photograph in conjunction with its ground I come to be in a relation to the object it represents.

With respect to item six, according to Peirce, there exist three kinds of pure ground: iconic, indexical, and symbolic [312, p. 150]. Icons, such as diagrams and images, are models grounded by their intrinsic (first order) similarity to the entity they represent. An index, such as a weathervane, signifies an entity because of a causal or spatiotemporal connection between the index and the entity. Symbolic representations, such as words, are grounded by convention. Symbols act as models only because of the way they are consistently interpreted, which can then generate regular effects on the behaviour of the agent.

In the final item, Von Eckardt interprets Peirce's theory of signs as naturalistic, meaning closely connected to natural science. The naturalistic approach fits neatly with the metaphysical assumption of supervenience outlined in Section 2.2.

I will now propose a definition of representation that reflects Peirce's triadic relation.

Definition 2.2 (Representation)**.** *A triadic relation between a model, entity and agent. The model substitutes for at least one entity, shaping the behaviour of at least one agent.*

The model *stands in* for an entity, and it always does so *for* an agent, thereby modifying the agent's predisposition to behave. In this definition, the interpretant is implicit in the ability of the model to shape the behaviour of an agent. The model may refer to a class of entities, and may also be shared by multiple agents. However, at least one entity and one agent are necessary for a representation relation.

2.4 External Representation

According to Peirce's triadic relation, the entity E is part of the world, the interpretant I is in the agent, but the location of the model M is unspecified. For the case that M is external to the agent, the triadic relation is relatively straightforward, since the problem of infinite regress does not need to be addressed. Peirce's typology describes three 'pure' types of grounding relations, which is important for a theory of representation in general. However, in practice, models may incorporate some combination of iconic, indexical and symbolic grounds. The aim of this section is to provide concrete examples of external models, and then show how entities, external models and interpretants have been classified within disparate academic disciplines.

2.4.1 Three kinds of external models

Common models used in representation can be distinguished by implementation rather than pure type. Here, I will assume that a model is somehow simpler than the entity it represents. Although not necessarily true, in practice this is reasonable, since a $1:1$ mapping in complete detail is in general completely useless (consider a life-sized map of the world, then consider trying to maintain the accuracy of every detail). Even if the representation bearer is not itself simple, practical models must confer some benefit, such as ease of manipulation[4]. Models have deliberate differences and may accentuate salient features, in order to retain only those aspects that are necessary to stand in for the entity. A caricature of a politician and a scale model of an aeroplane are examples of representations that can be understood and manipulated efficiently, which makes them useful substitutes for direct experience under certain conditions.

[4] An example of a useful $1:1$ mapping is the conversion between Polar and Cartesian coordinates, which is practical because performing this conversion can often improve the ease of manipulation.

One special kind of representation is a mathematical model. For example, two contained gas particles can be modelled mathematically by two hard uniform spheres with no internal energy except velocity, in an enclosed continuous four dimensional space (including time). The dynamics of the model constrain its behaviour by conserving momentum and energy, which is transferred along the axis joining the spheres' centres of mass when they collide elastically. The spheres are reflected by collisions with the containing walls. In principle, the model, in conjunction with initial measurements of position and velocity, can be used to predict the outcome of measuring the position and velocity of the particles at any future time. The model is a representation when someone (or more generally an agent) uses the model to stand in for a system of interest. For example, the agent could deduce the value of variables associated with the particles in place of direct observations of the gas particles at future times. In Peirce's typology, mathematical models have symbolic grounds. A mathematical model can always be interpreted as manipulating symbols in a formal system according to syntactic rules[5].

The gas particle dynamics can be represented in at least two other ways. Predictions could also be derived using a physical model, such as two balls on a billiard table. The billiard balls are analog representations, which are not arbitrary and abstract like symbols, but are in some way analogous to their subject. Formally, an analog model must exhibit systematic variation with its task domain [237]. This means the analog model does not have to represent every aspect of the entity in the same units – consider a sun dial, which represents the passage of time as the movement in space of a shadow. Animal testing of pharmaceuticals, architects' scale models and pictures are examples of analog representations, although note that the last two examples can also contain symbolic content, which is usually of a secondary nature. Note that in some analog models, it is possible to view the model from multiple perspectives, while other analog models may fix the perspective in the process of representing an entity. Either way, an agent must use the analog model in place of real world measurements in order to fulfil the representation relation. In Peirce's typology, analog models have iconic grounds.

Another way the gas particles could be represented is using the English language. For the task of predicting particle dynamics, language is quite limited. However, if the particles were at sufficiently low temperature that their movement was frozen, an English description of their configuration could provide a useful representation for an agent. Language can be used to arrange words, which function as labels, to represent objects. Nouns are labels for entities or classes of entities, while the verb phrase of a predicate with two arguments (two nouns) refers to the relationship between the corresponding entities. Labels in isolation can act as signs, which constitute the most primitive form of representation, capable of standing in for only a single idea. Signs are formalised in semiology, whose contemporary form follows much more closely from the work of Saussure [275] than Peirce [156]. When a set of signs is organised into a language with syntactic rules for manipulation and intricate networks of relationships between components, its representational power is qualitatively increased, and is rich enough to be studied in the distinct

[5]For some areas of mathematics, the corresponding formal system may have an infinite number of axioms, rules or symbols, however these areas of pure mathematics are not practical for forming representations, and for my purposes can safely be ignored.

but related fields of linguistics and structuralist philosophy. An important observation is that the structure of sign systems (languages) does not need to represent the structure of the world. The structure in language is based on the difference between terms, rather than a reflection of structure in the world. This decoupling both provides flexibility of expression within language, while at the same time necessarily limiting its representational nature. Following Saussure, this constructionist view of language is the dominant view in structuralist and post-structuralist philosophy. Note that while mathematics is also a language[6], and both mathematics and language have symbolic grounds, I consider formal systems separately from linguistic representations, because they can play significantly different roles in representation.

Many disciplines have developed explanations of the way external models – mathematical, analog and linguistic – are used by agents to represent their world. The three disciplines I now consider are metaphysics, military theory and systems theory. The terminology and the scope of representations under consideration varies significantly. In spite of this, it is found that Peirce's triadic relation provides a common structure for explaining representation in each case, and also that the external models conform to the three kinds identified in this subsection. Further, the theory of representation in general reveals shortcomings in each of the disciplinary accounts.

2.4.2 Representation in metaphysics

Popper advocated an ontological pluralist doctrine from 1967, which is detailed in *Objective Knowledge* [250] and concisely summarised in [251]. According to Popper, there exists three worlds:

- World 1 is the physical universe, including both living organisms and non-organic matter.
- World 2 is the world of individual psychology, of mental events, raw feels and thoughts.
- World 3 is the world of abstract products of the human mind, including language, scientific theories, mathematics, paintings and symphonies.

The use of 'world' is indicative of the ontological nature of Popper's distinction. He clearly views each world as consisting of different kinds of stuff, proposing "a view of the universe that recognizes at least three different but interacting sub-universes." [251, p. 143]. The nature of the interactions are causal, and the abstract world is always linked to the physical world via the human mind [251, p. 165]:

> If I am right that the physical world has been changed by the world 3 products *of the human mind*, acting through the intervention *of the human mind* then this means that the worlds 1, 2, and 3, can interact and, therefore, that none of them is causally closed.

[6] This interpretation is made precise in formal language theory.

Popper contrasts his three world hypothesis with ontological monism (materialism or physicalism) and ontological dualism (mind-body dualism) by saying that the monist only admits world 1, while the dualist only admits worlds 1 and 2. When Popper refers to say a symphony or a sculpture in world 3, this is separate from the world 1 instantiation of the entity. It is only the abstract ideal of the entity that exists in world 3. Thus, world 3 entities are types, which may have many corresponding real world tokens that are imperfect embodiments of their type. The key to Popper's defence of world 3 are the claims that a) abstract entities exist that are not embodied in world 1 or 2, such as the infinite members of the set of natural numbers $\mathbb{N}$; and b) abstract entities have a causal influence on world 1, such as Einstein's equation $e = mc^2$ resulting in the development and use of an atomic bomb.

The three world hypothesis is of interest to us, because it neatly separates the real world entities E that are being represented (world 1), mental interpretations I of those entities (world 2), and external models M that are products of the human mind (world 3), in a way that is compatible[7] with Peirce's triadic relation. However, Popper's cosmology directly contradicts our understanding of physics. In particular, conservation laws and symmetry imply that world 1 is closed, and current theory requires only four fundamental forces (the strong and weak nuclear forces, electromagnetism and gravity) to explain every causal physical interaction. Popper claims that world 3 entities that cannot be embodied in world 1 can nevertheless exert a causal influence on world 1, because they are apprehended by human minds in world 2, which then control causal events back in world 1. But then any physical explanation of a system that includes humans is causally incomplete. Even if one accounted for all of the interactions of the four fundamental forces, there would be a 'residual causality' that remained unaccounted for. This is because abstract entities are not subject to the four fundamental forces, and yet if they have their claimed causal powers, their absence or presence will change the aggregate force acting on bits of world 1 matter. Consequently, one can ask whether it is conceivable that an experiment exists that could test for a residual causality leak from world 1. This would require us to ascertain the presence or absence of an abstract entity, which would require a human mind, without affecting the physical state in the experiment. But in order to say whether the abstract entity was present, the memory would have to be stored in the brain, thus changing the physical state in the experiment (assuming supervenience). In fact, Popper's claim is metaphysical, and unfalsifiable, in contrast to the ideal of scientific conjecture that he advocated. For our purposes, Popper's ontological distinction is stronger than is justified. The same argument applies to the similar, but less sophisticated distinction that Penrose [245] proposes between the physical, Platonic forms, and the human mind.

[7]For both Peirce and Popper, M could be private or shared. However, there is a difference in emphasis: Peirce is mostly concerned with private use, whereas Popper concentrates on shared uses of M.

2.4.3 Representation in military theory

At the other extreme of the academic spectrum, one finds a position that as far as I can ascertain, is advocated only within the relatively isolated discipline of military theory. A central idea in Network Centric Warfare (NCW) [4] and the closely associated, but broader Effects Based Operations (EBO) [295] concepts, is that military actions occur in three domains: the physical, information and cognitive domains. They are based largely on "common sense" and are not rigorously defined. For example, Garstka [131] provides the circular definition: "The information domain is the domain where information lives." This definition is perpetuated in [4]. More sense can be made of Smith's [295, pp. 160-173] interpretation:

> The three domains provide a general framework for tracing what actually goes on in the stimulus and response process inside human minds and human organizations, and how physical actions in one domain get translated into psychological effects and then into a set of decisions in another domain. Understanding this process is important because with it, we can begin to comprehend how people and organizations perceive a stimulus or action and why they respond or react in the way they do and thus, how we might shape behavior.

Smith then defines each domain.

> ...the physical domain encompasses all the physical actions or stimuli that become the agents for the physical and psychological effects we seek to create. ...the actions in the physical domain may be political, economic, and/or military in nature, and all must be equally considered to be objects or events...
>
> The information domain includes all sensors that monitor physical actions and collect data. It also includes all the means of collating or contextualizing that data to create an information stream, and all the means of conveying, displaying, and disseminating that information. In essence, the information domain is the means by which a stimulus is recognized and conveyed to a human or to a human organization...
>
> The cognitive domain is the locus of the functions of perceiving, making sense of a situation, assessing alternatives, and deciding on a course of action. This process relies partially on conscious reasoning, the domain of reason, and partially upon sub-conscious mental models, the domain of belief. Both reason and belief are pre-conditioned by culture, education, and experience.

It is clear from these definitions that the physical domain contains the entities E that one needs to represent; the information domain is where external models M are displayed and disseminated; and the cognitive domain is where the models are made sense of – where the interpretants I exist. In both Smith [295] and Alberts *et al.* [4], the relationship between the domains is seen to be a flow from the physical domain to the cognitive domain via the information domain. Alberts *et al.* [4, pp. 12-13] establish this flow, and then use it to motivate the central importance of information:

> With the exception of direct sensory observation, all of our information about the world comes through and is affected by our interaction with the information domain. And it is through the information domain that we communicate with others (telepathy would be an exception). Consequently, it is increasingly the information domain that must be protected and defended to enable a force to generate combat power in the face of offensive actions taken by an adversary. And, in the all important battle for Information Superiority, the information domain is ground zero.

Disregarding the reference to telepathy and the sales speak, what is Alberts' claim? Direct sensation is claimed to be an exception to the usual flow of understanding from the physical world to the cognitive domain, via the information domain. This description, along with the accompanying diagram - reproduced in Figure 2.2, conjures up visions of the information domain as populated by automated sensors collecting, fusing and disseminating data unaided by human cognition and judgement. Smith is again more cautious, describing the cognitive domain as the locus where data is interpreted and decisions are made, not the information network. However, he maintains the same connections from the physical domain to the information domain, and the information domain to the cognitive domain. This is most explicit in the layered diagrams Smith uses to depict the domains, with the physical domain layer at the bottom, the information domain layer in the middle, and the cognitive domain layer on top.

NOTE: This figure is included on page 30 of the print copy of the thesis held in the University of Adelaide Library.

Figure 2.2: The three domains of warfare, after Alberts *et al.* [4].

Interestingly, the three domain model is derived[8] from Fuller's [129] book *The Foundations of the Science of War*, in which Fuller described a trinity between the three spheres of man. Because they are both based on a triadic relation, the

[8]In [131], Garstka notes that "A key element of the model is a focus on three domains: the physical domain, the cognitive domain, and the information domain. This conceptual model builds upon a construct proposed initially by J.F.C. Fuller in 1917, and refined in *Measuring the Effects of Network-Centric Warfare*."

structure of Fuller's three sphere theory of warfare is structurally analogous to Peirce's theory of signs. However, in the modern day interpretation of Fuller, the triadic relation between the domains of warfare has been reduced to two dyadic relations: the physical–information and information–cognitive relations. This intended "refinement" of Fuller's conceptual model has only concealed the essential nature of representation as an irreducible triadic relation.

2.4.4 Representation in systems theory

In his book on multidisciplinary thinking, Kline [185, p. 16] notes three uses of the word 'system' within science, which are relevant to the role of external representations. So that they can be compared, he gives them three separate labels. The first conception, the most common use by scientists outside the systems community, is:

Definition 2.3 (System (1))**.** *The object of study, what we want to discuss, define, analyse, think about, write about, and so forth.*

Kline refers to this understanding with the label '**system**', which for example could refer to the solar system, a communicating vessel, an ecosystem, or an operating system. In fact, according to Kline, a **system** can be anything, as long as there is a well defined boundary associated with the **system**. In this thesis, I use 'system of interest' to denote this meaning of system. In Peirce's triadic relation, the system of interest corresponds to the entity E.

The second usage is defined as:

Definition 2.4 (System (2))**.** *A picture, equation, mental image, conceptual model, word description, etc., which represents the entity we want to discuss, analyse, think about, write about, etc.*

Kline coins the term 'sysrep' to mean representations of systems. Sysreps are "one of three basic types of representation: words, pictures and mathematics": that is, sysreps are models M. The types of representations Kline identifies correspond to the categories of language, analog and mathematical models I proposed in Section 2.4.1, except that pictures are only one of several possible analogs. According to Kline, the ideal aim of a sysrep is to perfectly mirror a **system**, where "[b]y 'perfect mirror' we mean not only that the sysrep will fully represent each and every characteristic of the **system** with total accuracy, but also that it will represent nothing more" [185, p. 18]. The common – but misguided – conception of representation as a perfect mirror is critically examined in Section 2.5.2.

The third usage is the most general conception, which is consistent with attempts to define the meaning of system within the systems community:

Definition 2.5 (System (3))**.** *An integrated entity of heterogeneous parts which acts in a coordinated way.*

Kline uses the label 'system' or 'systemic' for this conception, where a systemic property is an emergent property, which is a property of the whole but not a property of the components of the system.

The final concept Kline invokes is a schemata, which denotes "all the ideas in a person's head which are used to represent and interact with the world" [185, p. 31]. Example schema include words, relational ideas, behavioural routines and medical diagnosis. Kline then answers the question: "What is the relation of a sysrep to schemata in the mind? A sysrep is a particular kind of schemata, a very special class of the totality of the schemata we construct in our minds." Kline defines non-mental representation as a special class and an extension of mental representation. This approach is problematic, because mental representation is actually more difficult to understand than external representation. It makes more sense to explain mental representation in terms of the more straightforward case of external representation, even if mental representation precedes external representation from a chronological view.

In Kline's view, the relation between entities, models and their interpretants is as follows. Scientists view the world as being comprised of **systems**, which are interpreted using mental schemata. Schemata enable complex interactions with the world, but are formed using largely non-conscious mechanisms, and may be fuzzy and unstructured. When we go to the trouble of making a schemata explicit and shared in a structured social environment, it becomes a sysrep (which must still be interpreted by people). The goal of forming sysreps is to mirror the **system**, so that ideally the behaviour of the sysrep and the **system** are identical. While Kline is right to distinguish between systems, system representations, and system interpretations, the details of how they interact are not consistent with Peirce's triadic relation.

Burke [72] has formalised and refined the systems approach to understanding representation, in a clearly articulated conceptual model. He offers the following definitions for entity, system, system description and model [72, pp. 9-12]:

Definition 2.6 (Entity)**.** *An entity is any object that has existence in the physical, conceptual or socio-cultural domains.*

Definition 2.7 (System (4))**.** *A system is an idealisation of an entity as a complex whole.*

Definition 2.8 (System description)**.** *A system description is a representation of a system.*

Definition 2.9 (Model)**.** *A model is an idealisation and/or representation of an entity.*

Four implications follow from these definitions. Firstly, because systems are idealisations of entities, they are abstractions that have no physical existence [81]. Systems are not part of the furniture of the world, they only exist inside minds. Stated another way, a system is a way of looking at the world [324]. Secondly, an entity can be idealised as a system in multiple ways: there is no unique systems view for any entity. Thirdly, and most importantly for this discussion on representation, both systems and system descriptions are considered to be models by Burke. The difference is that external models (a system description) presuppose the existence of a corresponding idealisation (a system). This is equivalent to requiring that external models M require an interpretant I in order to represent an entity E. Therefore, Burke's system theoretic interpretation is consistent with Peirce's triadic relation for external representation. Fourthly, Burke defines system descriptions to be derived from systems (idealisations of entities), rather than directly from entities. This implies that the system description can only capture aspects of the entity that have already been captured in the system. Consequently, a system description can be interpreted as a system that has been further abstracted from the entity it represents.

2.4.5 Summary of external representation

I will conclude the discussion of external representation by comparing the distinctions that have been identified above in different disciplines. The most notable commonality is that in each case, exactly three categories have been necessary to explain external representation, and furthermore these categories can be aligned with the entities, models and interpretants of Peirce's triadic relation. Of course, this has more to do with the selective nature of my literature survey than uniformity of approach. Descartes' [107] dualism was unconcerned with external representations, while Rosen's [265] Modeling Relation between the formal systems of science and the natural systems they represent attempted to explain external representations without explicit reference to the human mind or interpretants. Nevertheless, each of the approaches I have covered supposes that things are naturally considered as belonging either to the physical, the mental, or the social products of the mental. The physical world contains the entities that one would like to represent, external models are social products that can be shared, but they must be interpreted by someone or some agent to count as a representation.

There is an important way in which the domains in military theory differ from the accounts of representation by Popper, Kline and Burke. Interactions between the physical domain and the cognitive domain are mediated by the information domain. In contrast, the other accounts explain external models as products of the human mind. Physical entities must be conceptualised before they can be externally represented. Because militaries functionally separate the collection of information from decision-making, the role of human conceptualisation in information collection that mediates between the physical and information collection is easily ignored. But without human intervention and judgement there is only data, not representation or information, and automation can reduce but not eliminate human participation in constructing representations[9]. In view of Peirce's triadic relation, each of the alternative accounts considered aspects of this relation, but none are as comprehensive as Peirce's theory of signs.

External models have been variously held to be: abstract products of the human mind; information bearing artifacts; the socio-cultural environment; a specially precise subset of mental representation; a mirror that reflects part of the world; and a mental representation reduced by additional simplifying assumptions, which is explicit and shared. However, most of these assertions are not entirely accurate. Definitions, such as Kline's, that attempt to explain external representations with respect to mental representation are not enlightening, because the cause is more complicated than the effect. Peirce's typology of iconic, indexical and symbolic

[9] This is a point that Polanyi [249, p. 20] makes well, and an example is the automation of the photo-finish for horse races, which still required human judgement in a case where one horse was fractionally in front, but the other extended further past the finish line due to a thick long thread of saliva coming from the horse's mouth. It would seem that such semantic ambiguities cannot be satisfactorily resolved by syntactic processors.

pure forms, and my categories of formal, analog and linguistic models, provide a framework for understanding external representation, which is sufficiently general to account for representation across disparate disciplines. Within this framework, an external model is most accurately conceived of as a grounded representation bearer external to the agent who interprets the model. Less formally, an external model is an equation, analog or description that represents something for an agent and thereby modifies its behaviour.

2.5 Internal Representation

Representation plays an important explanatory role in biology. From the perspective of a living agent, the world contains limited essential resources of energy and matter for survival and reproduction, as well as threats to survival such as predators and other harmful energy sources. It is easy to see that the ability to sense qualities of the immediate environment and to control locomotion with context sensitive behaviour confers a significant relative selective advantage. Bacteria that follow a chemical or light gradient can be viewed as performing very basic representation: chemical reactions triggered by the local environment stand for greater expected concentrations of non-local useful energy which cannot be directly detected. An agent that can sense distal features of its environment, using passive or active sensors to detect patterns of incoming energy such as light photons or sound waves, can secure an even greater selective advantage. Whereas proximal sensory information requires an agent to 'bump' into a threat before it can react to it, an agent that can sense a threat at a distance can avoid the threat entirely. However, distal information is noisy, incomplete and intermittent. Just because a predator becomes occluded by vegetation does not secure the safety of its prey. Current sensory input alone is inadequate for determining the best action in any context. By constructing an internal representation of its environment, an agent can continue to act appropriately in the absence of direct sensory stimuli.

This story of representation in biology is inspired by the accounts of Dennett [105, pp. 177-182] and O'Brien and Opie [238], which suggest that representation is the problem that the brain is intended to solve. There is some empirical support for this conjecture in the form of the sea squirt *Ciona intestinalis*. The tadpole larva has a central nervous system of about 330 cells that controls locomotion. Once it attaches to a permanent object, it undergoes a metamorphosis that has been loosely described as eating its own brain (the cerebral ganglion is broken down and reused), since it no longer needs sensorimotor control, and therefore has no need to represent its environment.

Given this story, one may ask how the brain represents. This question has generated the most sophisticated conversation about internal representations, and has been especially preoccupied with the human brain. The Representational Theory of Mind, or representationalism, dates back at least to Aristotle [248]. The proposed answers of contemporary cognitive science divide into three main camps. They are Good Old Fashioned Artificial Intelligence (GOFAI), also known as symbolic, classical, or conventional cognitive science; connectionism; and the dynamical systems hypothesis.

When the representation is internal to the agent, one is faced with the question of what interprets the model. If internal models are interpreted in the same way as external models, then this leads to infinite regress, because the interpretant is also an internal model that requires its own interpretant, and so on. Von Eckardt [312, p. 283] describes two alternative resolutions to the infinite regress problem Peirce considered, and relates these to analogous moves in contemporary cognitive science.

The first solution is to weaken the definition of interpretant, to be a *potential* rather than an *actual* interpretant. The regress still consists of an infinite series of representations, but it is now easier to reconcile the associated interpretants, because they do not need to actually exist. This solution is reiterated by Cummins in cognitive science.

The second solution is "to find something that can function as an interpretant but which is not, itself, also representational and therefore in need of interpretation" [312, p. 283]. Peirce suggests that the only candidate for this is a habit-change. Specifically, Von Eckardt argues it must be a modification in the tendency to act in ways dependent on the content of the representation. The habit-change does not need to affect external behaviour; changes to mental habits (processes that generate other internal representations) also count. However, in order to eventually curtail the regress, internal models must ultimately be interpreted by shaping the agent's external behaviour. A very similar solution is suggested by Dennett, which Von Eckardt claims is the widely endorsed solution in cognitive science. Further, Von Eckardt [312, p. 290-302] shows in detail how this solution can handle the regress problem. Briefly, this involves demonstrating that:

- Interpretant I of model M is producible by M; and
- I is related to both the agent A and M, such that by means of I the content of M can make a difference to the internal states or the external behaviour of A towards the entity E.

Von Eckardt establishes this is the case for both conventional (symbolic) and connectionist machines. I will now provide a short introduction to GOFAI and connectionism, the two strongest advocates of representationalism.

2.5.1 Representationalism

> *How can a particular state or event in the brain represent one feature of the world rather than another? And whatever it is that makes some feature of the brain represent what it represents,* how does it come *to represent what it represents?*
>
> Daniel Dennett

These are the questions of representationalism, a position that assumes that the mind is a representational device, and that the brain has representational content. They are exceptionally difficult questions, because the mechanisms behind brain functions such as learning, memory and computation in the brain are currently poorly understood. Consequently, the mechanisms underlying representation are equally opaque. Also, under almost any metric, the human brain rates as the one of the most complex entities studied in science[10]. For a deeper discussion of representationalism than I can afford here, see Cummins [98].

As is the case for most enduring themes of Western philosophy, the first records of representational theories of mind are found in the writings of Aristotle [13]. In Book III, part 4, Aristotle describes the part of the human soul that thinks and judges: $\nu o\hat{\upsilon}\varsigma$ or the mind. According to Aristotle, the mind is "capable of receiving the form of an object; that is, must be potentially identical in character with its object without being the object." This statement clearly demonstrates Aristotle's use of the distinction between a model and the entity it represents. By form, Aristotle refers to the properties of the object, as opposed to its material substance. In Aristotle's metaphysics, the immaterial mind knows something when it takes on the form of that object, such that it represents the object in virtue of their similarity, in exactly the same way that a picture can represent a scene (Peirce's iconic grounds). "To the thinking soul images serve as if they were contents of perception …That is why the soul never thinks without an image." Berkeley [44] and Hume [163, 164] both extended this Aristotelian conception to argue that all mental contents are images in the mind, and that they are representations in virtue of their resemblance to perception. The inherent weakness of basing mental content on similarity can be seen by probing the mechanisms that could imbue mental images with the same properties as the objects they represent. Images presented to an immaterial mind are not so much an explanation as a metaphor, where thinking is like putting on a theatre for the Eye of the Mind.

In contrast, Hobbes [151, Chapter V] and Leibniz [202] advanced the idea that everything done by the mind is a computation. In this view, thought proceeded by symbolic manipulations analogous to the additions and subtractions of the new calculating devices – in modern parlance the mind was seen as an "automatic formal system" [149]. Notably, this reframed the question of representationalism to propose a mechanical and material explanation of mental processes. This provided a crucial step towards a science of cognition, because it opens up the possibility that certain features of cognition could be reproduced artificially.

The link between computation and representation is important but subtle. Because of the universality of Turing's conceptual model of digital (symbolic) computation, it is a common assumption that all computation is equivalent to a Universal Turing Machine. However, as O'Brien and Opie [239] correctly point out, this does not account for analog computation. They propose a definition of computation in general, which is broad enough to capture both analog and digital computation, but still sufficiently constrained to differentiate computation from the vast majority of physical systems – intestines, microwave ovens, cups of tea, etc. – that are not involved in computation.

> [T]here are two distinctive features of computational processes (as opposed to causal processes in general). First, they are associated with representing vehicles of some kind. Second, and more importantly, computational processes are shaped by the contents of the very representations they implicate. We thus arrive at the following characterisa-

[10] See Bar-Yam [25] for estimates of the complexity of the brain compared with other systems.

tion:

> *Computations are causal processes that implicate one or more representing vehicles, such that their trajectory is shaped by the representational contents of those vehicles.*

This characterisation of computation makes explicit the link between computation and representation. Computations are those processes involving representing vehicles (models), such that the outcome of the process depends on the content of the model. Representation is inherent in computational processes, and computation is the mechanism that causally links the contents of models to changes in the behaviour of the agent that interprets the model. Representation and computation are a package deal: a commitment to a computational theory of mind entails a commitment to representationalism.

Although conceived in the 17$^{\text{th}}$ century, it was not until the mid 20$^{\text{th}}$ century that the computational idea rose to prominence. The initial hype associated with the AI movement had a profound impact on 20$^{\text{th}}$ century cognitive science, such that computational theories of mind were predominantly based on algorithmic symbol manipulation. The Universal Turing Machine [301] provided a theoretical basis for universal symbol-based simulators of human intelligence, while exponential increases in computing power dramatically expanded the application of computer algorithms towards focussed engineering tasks that had previously required the application of the human mind.

Yet simulations that could be confused with intelligent humans have not materialised. AI researchers began to hit some fundamental walls: general intelligence appeared to require fast, situated, unencapsulated reasoning, where automated formal systems were slow, abstract, and only capable of manipulating the initial axioms they were given according to fixed rules. Coinciding with a growing dissatisfaction with the ability of the products of AI to live up to expectations, several alternatives have been advanced within cognitive science, and symbolic computational models of cognition began to be referred to as GOFAI.

Connectionism, or Parallel Distributed Processing (see for example [218]), which is based on highly abstract networks of artificial neurons, presents an alternative paradigm for modelling cognition, which can be interpreted as performing analog computation. Connectionist models are inspired by current understanding of the architecture of the brain, and are described by Dennett [105, p. 269] as blazing the first remotely plausible trails of unification between the mind sciences and the brain sciences. Different kinds of connectionist networks have been shown to have content addressable memory [159]; provide universal function approximation [160]; degrade gradually when damaged; and distribute representations across the set of connection weights, which decouples representations from individual symbols. Due to their parallel processing, connectionist networks are also very fast. The theoretical model of connectionism, an artificial neural network with real connection weights, has been proven to be capable of hypercomputation [291] – that is, able to compute functions that Turing machines cannot. Of course, such machines are not practical, since real numbers in general require infinite information, and there are also a number of issues artificial neural network implementations suffer from. They are almost always simulated on a digital computer, which implies these instantiations are equivalent to Turing machines; they learn reliably only under supervision; and they are usually treated as black boxes, because their behaviour is not currently well understood. Of course, there are philosophical concerns as well. For example, Fodor and Pylyshyn [126] criticise connectionism because it cannot explain systematicity: the feature of human cognition whereby the ability to think one thought entails the ability to think of numerous logically related thoughts, such as its converse.

The most recent alternative to both GOFAI and connectionism is the dynamical systems hypothesis [304]. However, advocates of the dynamical systems account are often explicitly critical of explanations involving representation, so a discussion of dynamical systems is deferred to Section 2.5.2 on anti-representationalism.

In summary, representational theories of mind have been proposed that are based on symbolic manipulation and analog covariance. GOFAI and connectionism agree that mental contents can stand in for, and stand in relation to real world objects. They also assume that psychological processes are computations that represent aspects of the external world.

2.5.2 Anti-representationalism

> *There is no harm in saying of good tools and good moves that they are also good representations, but nothing interesting is conveyed by this choice of idiom, and its employment should not tempt us to construct theories about how representation works.*
>
> Richard Rorty

Accounts of representation in cognitive science and artificial intelligence have been criticised as a basis for biological behaviour on a number of fronts. Brooks [64] summarises one key idea against representation in the physical grounding hypothesis: the world is its own best model, the trick is to sense it appropriately and often enough. The first part is true but uninteresting, because it is the trivial case where the representation relation degenerates into a diadic relation between A and $E \equiv M$. The second part is important, because it emphasises the need for agent decisions to be grounded. However, the physical grounding hypothesis does not and can not dispose of representation entirely. Even if the world is used as its own model, the agent needs to interpret the meaning of its observations. Constructionist accounts of vision (see for example [243]) argue that the process of perception involves significant construction by the observer. In their critique of a simplistic but common conception of "pure vision" – essentially the idea that the visual system is a bottom-up hierarchy designed to fully mirror the visual scene – Churchland *et al.* [82] provide an alternative account that they label interactive vision. Some of the constructive characteristics of interactive vision are: visual fields are highly non-uniform; vision is exploratory and predictive; the motor system and the visual system are entangled; sensory processing is more like a recurrent network than a hierarchy; and vision cannot be neatly separated from other brain functions. Consequently, the process of sensing the world appropriately is in fact one of the major sources of representational activity in the brain [241]. Also, in

Section 2.3 I gave a number of reasons why internal representations can be convenient, even if they are not perfect substitutes for unmediated access to the real world.

Van Gelder [304] denies that cognition involves computation or representation, advancing an alternative dynamical systems hypothesis. In this account, rather than interpreting cognitive states as symbols, they are treated as quantifiable states of a nonlinear dynamical system. Van Gelder uses the Watt governor, depicted in Figure 2.3, to illustrate his thesis. The Watt governor is a mechanical device that maintains a constant speed for a flywheel despite fluctuations in both steam pressure from the boilers and the engine workload. The Watt governor was a pivotal invention during the industrial revolution that allowed the generation of reliable, smooth and uniform power. The Watt governor works because a spindle is geared into the flywheel such that the spindle rotates proportionally to the speed of the flywheel. The faster the spindle rotates, the more centrifugal force it generates, raising the spindle of the flywheel. Because the spindle is directly linked to the throttle valve, the faster the spindle rotates, the higher its arms rise, the more the valve is closed, restricting the flow of steam. As the speed of the flywheel decreases, so too does the spindle, the arms fall, opening the valve and increasing the flow of steam. Thus, a steady state for the speed of the flywheel exists and the Watt governor maintains the steady state by exerting negative feedback on any deviation from the steady state.

NOTE: This figure is included on page 39 of the print copy of the thesis held in the University of Adelaide Library.

Figure 2.3: The Watt centrifugal governor for controlling the speed of a steam engine, after [121] as reproduced in [304].

Van Gelder compares this mechanical device, which he classifes as a dynamical system, with a hypothetical computational device capable of performing the same function. The computational device would follow an algorithm that depends on representation.

> The very first thing it does is measure its environment (the engine) to obtain a symbolic representation of current engine speed. It then performs a series of operations on this and other representations, resulting in an output representation, a symbolic specification of the alteration to be made in the throttle valve; this representation then causes the valve adjusting mechanism to make the corresponding change [304, p. 350].

In contrast, the mechanical device is non-representational. Van Gelder gives four reasons: representation is not needed to fully account for the operation of the Watt governor; the obvious correlation between arm angle and engine speed is not representational because representation is more than mere correlation; the simple correlation only obtains in the steady state; and the arm angle cannot represent engine speed because the two quantities are coupled.

Of these, the first three reasons are not persuasive. Just because an explanation of the Watt governor within some frameworks do not need the concept of representation does not imply that representation *cannot* be used to explain the same process. After all, none of the compound objects – such as spindles and throttles – are necessary concepts in the quantum mechanical wavefunction of a Watt governor. The second point does nothing to disprove representation occurs, it merely demands a higher standard of proof than demonstrating correlation, while the third point only notes that any correlation is not simple.

The forth reason is the most interesting. Van Gelder observes that "the angle of the arms is at all times determining the amount of steam entering the piston, and hence at all times both determined by, and determining, each other's behaviour." Because of this circular causality, Van Gelder claims that representation is "the wrong sort of conceptual tool to apply". When representation is thought of as a mirror, it does indeed seem wrong for the mirror to determine any part of the mirrored entity, because there is an asymmetry in their relationship. However, under the conception of representation as a triadic relation, it is *necessary* for the model to change the behaviour of an agent, and *possible* for the agent to be acting upon the represented entity. Peirce's triadic relation does not preclude the formation of feedback loops, although it does provide an incomplete explanation for such tightly coupled variables as the arm angle and engine speed.

The important criticisms of both Brooks and van Gelder are directed at the cognitive science community's early preoccupation with explicit symbolic representation. However, Section 2.3 demonstrates that representation in general can have iconic and indexical – not just symbolic – grounds. Brooks' situated robots do not do away with representation altogether – they actually encode significant amounts of their behaviour symbolically on finite state machines. Dynamical systems, as advocated by Van Gelder, can still function as representations with iconic grounds. The analog model in Section 2.4.1 is one such example. Rather than undermining representation, these critiques serve to highlight differences between formal systems and other possible bases for biological representation.

Maturana and Varela's [217] ground-breaking second order cybernetics approach to the biological basis of cognition is also critical of representationalism, which they claim is inadequate for a scientific explanation. They use an analogy reminiscent of Searle's [277] Chinese room argument to claim that living systems do not represent [217, p. 136]:

> Imagine a person who has always lived in a submarine. He has never left it and has been trained how to handle it. Now, we are standing on the shore and see the submarine gracefully surfacing. We then get on the radio and tell the navigator inside: 'Congratulations! You avoided the reefs and surfaced beautifully. You really know how to handle a submarine.' The navigator in the submarine, however, is perplexed: 'What's this about reefs and surfacing? All I did was push some levers and turn knobs and make certain relationships between indicators as I operated the levers and knobs. It was all done in a prescribed sequence which I'm used to. I didn't do any special maneuver, and on top of that, you talk to me about a submarine. You must be kidding!'

This analogy works by specifying an overly narrow system boundary. The adequacy of the navigator in avoiding the reefs cannot be explained unless the boundary is expanded to include the process that generated the prescribed sequence of knob turns and lever pushes. Specifically, the person in this example only becomes a navigator once they have been *trained*. But then it is easy to see that the precise purpose of training the navigator in the sequence of actions is *to stand in for* observations of the reefs and the depth below sea level, and thereby modify the submarine's behaviour.

A more serious threat to representationalism is anti-representationalism, which has been advocated by Davidson, and even more forcefully by Rorty [263]. Anti-representationalism holds that any statement about the world is an inseparable cohabitation of subject and object, rather than correspondence between an object and a model. Rorty rejects the 'mirror' metaphor of knowledge, where knowledge is a reflection of the mind-external world. According to Rorty this metaphor, which we have already seen used explicitly by Kline above, is the central metaphor for representationalism. Rorty criticises what he calls the Aristotle-Locke analogy of knowledge to perception,

> ...the original dominating metaphor as being that of having our beliefs determined by being brought face-to-face with the object of the belief (the geometrical figure which proves the theorem, for example). The next stage is to think that to understand how to know better is to understand how to improve the activity of a quasi-visual faculty, the Mirror of Nature, and thus to think of knowledge as an assemblage of accurate representations. Then comes the idea that the way to have accurate representations is to find, within the Mirror, a special privileged class of representations so compelling that their accuracy cannot be doubted. These privileged foundations will be the foundations of knowledge, and the discipline which directs us toward them—the theory of knowledge—will be the foundation of culture. The theory of knowledge will be the search for that which compels the mind to belief

> as soon as it is unveiled. Philosophy-as-epistemology will be the search for the immutable structures within which knowledge, life, and culture must be contained—structures set by the privileged representations which it studies. The neo-Kantian consensus thus appears as the end-product of an original wish to substitute *confrontation* for *conversation* as the determinant of our belief [263, p. 163].

It should be noted that Rorty's attacks are not directly focussed on the cognitive science debate on mental contents, which the previous critiques have participated in. Rorty, who is trained in analytic philosophy, is more concerned with structuralist and linguistic attempts to ground knowledge as representation. For example, Rorty [264] cites Brandom's [60] characterisation of the representationalist school as saying that "the essential feature of language is its capacity to represent the way things are." Proponents of this school are taken to be Frege, Russell, Tarski and Carnap, who are contrasted with Dewey, Wittgenstein and Sellars, who view language as a set of social practises. However, it was noted in Section 2.4.1 that there is no necessity for language to be representational. Further, Rorty's target is much larger than just the philosophy of mind – he seeks to question the legitimacy of transcendental (Kantian) epistemology as distinct from psychology, and advocates that the only constraints on knowledge are essentially conversational in nature. Rorty rejects the very idea of a theory of knowledge, truth or rationality.

Is it possible to salvage representationalism from these attacks? Rorty seeks an *a priori* defeat of representation, but there is an empirical component to the question of whether minds represent. Languages are not reflections of reality, in the sense that they are not mirrors that can be polished to provide a True representation of the world. However, it does not immediately follow that formal systems, analogs or minds can not represent in any meaningful way.

A stronger defence of representationalism can be made by carefully articulating what work the representation relation needs to perform. From the perspective of an agent faced with a decision, consider two different processes for choosing between the alternatives. For concreteness, suppose the agent is a frog near the edge of a cliff choosing whether to jump forward to the left or right. One choice is safe but the other choice will result in certain death. Using the first process, suppose that the frog, like one of Brooks' situated robots, is able to make its decision by using the world "as its own best model". Brooks' idea of sensing the world often enough is similar to Van Gelder's claim that dynamical coupling and feedback can do the same job as representation, without requiring extensive planning or computation. In the case of the frog, it is a pretty simple and efficient process: both alternatives are sensed and the apparently less perilous alternative is immediately acted upon. The process is memoryless[11], so it can be repeated every time the frog lands. It can work when several conditions are met: if the frog can sense at least as far as it leaps; if the sensory comparison is reliable; and if the environment is sufficiently stationary while the frog is airborne.

If the latter condition is violated, nothing can guarantee the safety of the frog if it

[11] I should clarify that I mean memoryless in the mathematical (Markovian) sense: the probability of future states only depends on the current state. This is significantly more abstract and general than the meaning of memory in cognitive science.

continues to leap. However, by using an alternative representational process, the frog may stand a chance even when the first two conditions do not hold. Suppose that it is a dark and foggy night, so that the frog cannot reliably sense the relative merits of jumping left and right. Fortunately, the frog has taken this path many times before, and remembers the sequence of left and right jumps that have got it home safely in the past. Then recognition of the starting location and recollection of this sequence can substitute in the decision-making process for sensing the alternatives at every step. The sequence can act as a model which when interpreted by the frog can stand in for the currently unreliable sensory information. In this process, representation must still be grounded by having previously sensed the alternatives using the first process. However, it can no longer be memoryless. By maintaining an internal model, the frog amplifies the applicability of sensory information beyond immediate local sense-response reflexes, to affect behaviour non-locally in space or time. Representation allows the agent to do more with less sensory information, to fill in gaps and to generalise new information. In other words, the real work that representation is doing is inference.

Rorty is right to reject the metaphor of representation as a mirror, reflecting the nature of reality for Descartes' Eye of the Mind. But representation is not a mirror, its purpose is not to *reflect* but to *infer*. In this section, I have argued that perception is to some extent constructed, which involves representational activity. I have shown that representation can have non-symbolic grounds, meaning dynamical systems can be a basis for representation. I have examined Maturana and Varela's argument against representation, which relies on an overly narrow definition of system boundary. Finally, I have argued that Rorty's attack on representation is largely directed towards language as representational, and to representation as mirroring. These critiques have merit, but do not challenge the general theory of representation provided by Peirce's triadic relation as an explanation of internal representation.

2.6 Agents

So far I have not specified what I mean by an agent. However, the preceding discussion on representation offers a precise way of characterising agency. Because of the ubiquity of agents in complex systems, this section contains the most important implications of this chapter.

As Peirce has argued, representations require an interpretant, and therefore an agent to perform the interpretation. Thus, there is a sense in which representations (and computations) are relative to a subject – that is, they are subjective [278, p. 92]. But there is an equally objective way that a subject plus a model either does, or does not, represent. By redrawing the system boundary to include the model's user, representation is an intrinsic feature of this system of interest. This is the importance of the triadic relation: because it is irreducible, the system boundary must always extend to include the agent in order to understand representation.

This is why, unless one can identify who is using the model and how it shapes their behaviour, the model cannot be considered to be a representation. Peirce, Von Eckardt, Popper, Smith, Kline and Burke all identify the 'who' with a human mind, which is more generally the case in the literature. I have generalised this to say that a model is always a model employed by an agent. An agent can be a person, but it can also be a group of people, an animal, a cell, a certain kind of robot or a certain kind of physical process (after all, each member of this list is a physical process). The ability to form and use representations appears to be the principle difference[12] between an object and an agent – it is the difference between kicking a rock and kicking a cat (not that either experiment is condoned, except as a thought experiment).

More formally, if an entity's response to a stimulus is directly determined by its current state, and the current state does not include any models, then the entity is not an agent. If stimulus and response are indirectly related because they are mediated by representation, then the entity is an agent.

Definition 2.10 (Agent)**.** *An entity that constructs and uses representations to shape its own goal-directed behaviour.*

More will be said about goal-directed behaviour below, but for now note that goal-directed behaviour does not imply that agents only have a single goal: it is merely intended to distinguish between directed and undirected behaviour. It seems there is a continuum, such that entities may have a degree of agency, depending on how indirect the relationship between stimulus and response is, and how sophisticated the representations can become, which is often called the plasticity of the representing medium. I am not overly concerned about the precise demarcation between agent and non-agent. The definition is more useful for comparative purposes, in order to investigate if the degree of agency has increased, and to say that a human has more agency than a cockroach, which has more agency than a virus, which has more agency than the robot Cog [66], which has more agency than a cyclone[13].

The degree of autonomy of an agent refers to freedom of choice or variety, which is made precise by the notion of source coding in information theory (see Section 5.2). The degree of autonomy is evident in the sensitivity of changes in the behaviour of the agent to changes in its representations. For example, if an agent's model is replaced by any other model (such as its inverse) and yet this has no causal influence on the behaviour of the agent, then the model does not contribute to the autonomy of the agent. If this holds for all models, then the agent is not autonomous. A model must shape behaviour to be a representation and provide the agent with autonomy. In contrast, if any arbitrary desired feasible state within the agent's environment can be achieved by changing only the agent's representations, which then realise the desired state by modifying the agent's behaviour, the agent has maximum autonomy. When autonomy is shared between two or more agents,

[12]For example, Aristotle [13] says "The soul of animals is characterized by two faculties, (a) the faculty of discrimination which is the work of thought and sense, and (b) the faculty of originating local movement." In my account, representations both encode distinctions and shape movement or behaviour.

[13]Most people do not consider a cyclone to be an agent, even though it is a self-maintaining, non-equilibrium entity with unpredictable behaviour. The anti-cyclonic Great Red Spot on Jupiter is a structure with a diameter significantly greater than Earth, which has persisted since it was observed by Cassini in 1665. I believe the main reason cyclones are not considered agents is because it is not possible to sustain an interpretation of either goal-directed behaviour or representation for a cyclone and although unpredictable, they are not autonomous.

this is the subject of game theory, and the degree of autonomy of a player is the number of available strategies[14], and any mixed strategy on this set constitutes a model.

The autonomy of agents can lead to philosophical debate about free will and teleology. In view of Hume's compatibilism [163, 164], the autonomy of an agent does not imply the agent is necessarily nondeterministic – that with *exactly* the same internal and external states, two distinct responses to the same stimulus are possible. Instead, a weaker condition holds, namely given different representations, an agent is capable of choosing different actions. To confuse the matter, often it is useful to explain the behaviour of a system as an autonomous agent, even when it is clearly not purposive. For example, Dawkins [101] describes genes as selfish molecules, as if they have minds, which is a form of teleonomic explanation. For my purposes, I will always assume that agency entails a degree of autonomy in Hume's sense, and also implies that the agent is capable of exhibiting goal-directed behaviour.

The notion of goal-directed behaviour has been formalised by Sommerhoff [297], who observed that the essence of goal-directed activity is:

> that the occurrence of the goal-event G is *invariant* in respect of certain initial state variables ($\mathbf{u}_0$) *despite* the fact that G depends on action factors and environment factors that are *not* invariant in respect of $\mathbf{u}_0$. The invariance of G being due to the fact that the transitional effects of changes in $\mathbf{u}_0$ mutually compensate, so to speak.

Sommerhoff realised it was possible to treat goal-directedness as an objective property, independently of the subjective notion of purposiveness of interest to the psychologist. He established three necessary and sufficient criteria for goal-directed behaviour. Firstly, for at least one variable $\mathbf{a}$ associated with the action, and one variable $\mathbf{e}$ associated with the environment, for at least one time t_k,

$$F(\mathbf{a}_k, \mathbf{e}_k) = 0. \tag{2.1}$$

This ensures that the action is capable of compensating for environmental variability. Secondly, $\mathbf{a}$ and $\mathbf{e}$ must be mutually orthogonal, meaning that the value of one of the variables does not determine the value of the other for the same instant. This allows the mechanism for goal-directed behaviour to realise Equation (2.1) for a range of initial conditions. And thirdly, there must be a set S_0 of initial environmental conditions (where $|S_0| \geq 2$), such that each initial condition requires a unique action $\mathbf{a}_k$ which satisfies Equation (2.1). This criterion ensures that the goal could have been achieved from an ensemble of initial conditions, rather than only from the actual initial conditions. $|S_0|$ provides a measure of the degree of goal-directed behaviour: the greater $|S_0|$ is, the more environmental variety the agent can destroy and still achieve its goal. Thus, goal-directed behaviour is underpinned by Ashby's [18, 19] law of requisite variety, which is explained in Section 3.3.2.

From a stimulus-response perspective, an agent can be thought of as sensing stimuli and acting to produce an appropriate response. The function that maps from

[14]This game theoretic interpretation assumes the set of strategies is countable, where only strategies that affect the value of the game are counted.

sensory inputs and models to output actions is its decision map. The sense, decide and act functions of agents are roughly analogous to detectors, rules and effectors in Holland's [153] complex adaptive systems terminology; perceptual, cognitive and motor components in cognitive science; input (actuating signal), control unit (dynamic element), and output (controlled variable) in control theory; state, policy and action in reinforcement learning [298]; stimulus, organism and response in Hull's [162] version of behavioral psychology; state, mixed strategy and move in game theory; and observe, orient/decide and act (OODA) in Boyd's [57] decision cycle.

The sense, decide, act trinity is a pervasive characterisation of agency that can be related back to Peirce's triadic relation. Without sensing, there is no way to ground representations. Without acting, representation cannot shape external behaviour. Without deciding, representations cannot be interpreted, the agent cannot be autonomous, and its behaviour is not goal-directed. A necessary and sufficient condition for agency is the possession of sense, decide, and act functions. But this is exactly equivalent to requiring that an agent be able to construct and use representations to shape goal-directed behaviour.

In summary, representation and agency have been co-defined. Representations always involve an agent, and agents always represent their environment. The triadic nature of the representation relation is the reason that these definitions cannot be separated. Due to this intimate relationship, a theory of representation is essential to an understanding of the behaviour of agents in a complex system.

Chapter 3

Systems

What is a systems approach? The first step towards answering this question is an understanding of the history of the systems movement, which includes a survey of contemporary systems discourse. In particular, I examine how systems researchers differentiated their contribution from mechanistic science – but also from holistic doctrines; and identify the similarities and sharpest differences between complex systems and other systems approaches. Having set the scene, the second step involves developing a definition of system for this thesis, consistent with the spirit of the systems approach.

3.1 Introduction

> *If someone were to analyse current notions and fashionable catchwords, they would find 'systems' high on the list. The concept has pervaded all fields of science and penetrated into popular thinking, jargon and mass media.*
>
> Ludwig von Bertalanffy, 1968

Since von Bertalanffy's [308] theory of open systems introduced the idea of a General Systems Theory (GST) which rose to prominence in the mid twentieth century, the field of systems research has become ever more fashionable, and the closely knit couple of GST and cybernetics have given birth to a large family of systems approaches, including complex systems [25, 132, 175], nonlinear dynamical systems [140], synergetics [142], systems engineering [143], systems analysis [110], systems dynamics [127], soft systems methodology [81], second order cybernetics [217, 314], purposeful systems [2], critical systems thinking [168, 169, 303], total systems intervention [125], and systemic therapy [37]. As well as adding systems concepts to the tool sets of all fields of science, the systems approach has opened up new areas within disciplines, such as systems biology [183], and created new hybrid disciplines at the interface between traditional disciplines, such as sociobiology [330]. Many of these systems approaches are introduced in Midgley's [222] epic four volume collection on systems thinking. Diversity is clearly a major strength of the systems approach, but this also makes it difficult to characterise. Consider the following typical definition of a system:

Definition 3.1 (System (5))**.** *A system is a set of entities with relations between them [195].*

By this definition, the converse of a system is a set of entities with no relations, not even logical ones, between them: a heap[1]. But heaps cannot exist physically, they are only an idealisation, since logical relations can always be established between a set of entities. Alternatively, it could be argued that an indivisible 'atomic' element does not meet the definition (provided trivial sets and relations are excluded). Thus a system is never undifferentiated. Nevertheless, under this definition, just about everything bigger than an electron is a system, which makes 'system' a vacuous container concept until it is further qualified. For example, the following systems engineering definition only considers physical systems in functional relationships:

Definition 3.2 (System (6))**.** *A system is a bounded region in space-time, in which the component parts are associated in functional relationships [50].*

While there have been many attempts consistent with this definition to make systems the Furniture of the World[2] [68, vol. III], claims that the world is systemic – usually synonymous with naïve realist and 'hard' systems research – are problematic. They ignore the fact that systems thinking involves its own set of simplifying assumptions, modelling choices and reductions. They also ignore the insights of second order cybernetics [314], that the observer (and by extension the systems theorist) should also be considered as a system. However, there is no need for systems approaches to make the bold claim that the world is "made up of" sys-

[1] Aristotle [14] used the concept of a heap to refer to matter without form.

[2] That is, give systems an ontological status or "real" existence.

tems. Knowledge that is obtained using a systems approach makes more sense when it is seen as one perspective for thinking about the world, rather than an objective property of bounded regions of space-time. Consequently, a definition of system should define how the real world is idealised, represented and acted upon when viewed as a system. A definition of system will not tell us how to discern when a part of the world "is a system", but it can shed light on when it may be appropriate to utilise a systems approach.

It is also important to stress that the opposite extreme of a relativist and purely subjective account of systems, where all views are equally valid, is not appropriate. In this chapter, I will take a moderate approach that considers the insights from both hard and soft systems approaches. The utility of a systems perspective is the ability to conduct analysis without reducing the study of a system to the study of its parts in isolation. Knowledge obtained from the systems perspective can be as "objective" as knowledge obtained from a reductionist scientific perspective. In this way, a systems approach contributes valuable insight capable of complementing the traditional analytic method.

This chapter begins by describing the scientific climate prior to the systems movement. Next, the history of systems thinking is interpreted. There are many different stories that could be told to make sense of the development of the systems approach[3]. This history is selectively biased towards the design of complex systems as the theme of my thesis. It emphasises the key concepts that differentiate between members of the systems family. The systems approaches I survey are general systems theory, cybernetics, systems analysis, systems engineering, soft systems and complex systems. The introduction of specific systems results is mostly deferred to Chapter 5. Finally, I develop a definition of 'system' as a concise summary of the systems approach.

3.1.1 Novel contributions

I am the sole author of this chapter, which has not been published elsewhere. My original contributions are:

1. My history of the systems movement;
2. A comparison between the research agenda in general systems theory and cybernetics with complex systems;
3. A list of features of the systems approach;
4. A definition of system as a representation of an entity as a complex whole open to feedback with its environment;
5. A list of key assumptions that should be critically examined as part of a systems approach.

[3]Alternatives include Checkland [81] and Matthews [216], or see Lilienfeld [208] for a particularly spiteful critique.

3.2 Science Before Systems

If the Renaissance was a period of re-discovery of classical Greek science, then the subsequent period of the Enlightenment[4] produced the scientific revolution that provided a foundation for the Modern worldview [216]. Descartes, the 'father of modern philosophy', played a pivotal role in the self-understanding of Enlightenment science. In an attempt to demarcate between knowledge derived from science/philosophy and superstition, Descartes described a scientific method, the adherence to which he hoped could provide privileged access to truth. Descartes' [108] method contained four precepts:

> The first was never to accept anything for true which I did not clearly know to be such; that is to say, carefully to avoid precipitancy and prejudice, and to comprise nothing more in my judgement than what was presented to my mind so clearly and distinctly as to exclude all ground of doubt.
>
> The second, to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution.
>
> The third, to conduct my thoughts in such order that, by commencing with objects the simplest and easiest to know, I might ascend by little and little, and, as it were, step by step, to the knowledge of the more complex; assigning in thought a certain order even to those objects which in their own nature do not stand in a relation of antecedence and sequence.
>
> And the last, in every case to make enumerations so complete, and reviews so general, that I might be assured that nothing was omitted.

The first rule of sceptical inquiry, and the fourth rule of broad and complete analysis, describe the dominant approach to Modern Western philosophy[5]. Meanwhile, the second rule of analytic reduction, and the third rule of understanding the simplest objects and phenomena first, became influential principles of Modern science, which would eventually differentiate itself from philosophy. Together, these two principles provided the view of scientific explanation as decomposing the problem into simple parts to be considered individually, which could then be re-assembled to yield an understanding of the integrated whole. It is this that I will refer to as the Cartesian analytic method.

While Descartes had articulated a philosophy for the scientific method, Newton's [234] breakthroughs in gravitation and the laws of motion showed the immense power of simple, precise mathematical idealisations to unify and at the same time

[4] The Enlightenment usually refers to the period between the signing of the peace accord at Westphalia in 1648, which brought stability to Western Europe, and the publication of Kant's [173] *Critique of Pure Reason* in 1781, which mounted a sceptical challenge to the Enlightenment philosophy.

[5] The first rule is hardly a contentious principle in philosophy, but the fourth rule is equally important. For instance, Broad [62, p. 12] claims that "the greatest mistake in philosophy is that of over-simplifying the facts to be explained."

quantitatively predict the behaviour of diverse phenomena. Newton's Law of Universal Gravitation

$$F = \frac{Gm_1 m_2}{r^2} \tag{3.1}$$

gives the force of attraction F between two objects idealised as point masses m_1 and m_2, where r is the distance between them and G is a universal constant. The law describes the mechanics of gravity by a universal fixed rule, which captures the first-order effects that dominate the dynamics of most macroscopic inanimate bodies. In particular, our solar system is dominated in terms of mass by the Sun, and the inverse square relationship of gravitational force to distance means that other solar systems exert insignificant forces on our own. This allowed Newton to ignore the great majority of objects in his calculations of planetary motion – of the 10^5 objects in our solar system, Newton only had to consider 10 [324, p. 13]. Further, by Equation (3.1), the force between m_1 and m_2 is independent of any other mass m_i. This allows each pair of interactions to be considered independently and then summed: superpositionality holds for the Law of Universal Gravitation, allowing tremendous simplification. Not only did Newton's mechanics provide an alternative cosmology that ended the dominance of Aristotle's worldview, it unified terrestrial and celestial dynamics. Once the heavens and Earth were seen to be governed by the same laws, it became conceivable that the mathematics that accounted for the motion of planets could also account for life on Earth.

Newton's mechanics played a primary role in the 19$^{\text{th}}$ century physics worldview of a deterministic, mechanistic Universe, which von Bertalanffy [309] characterised as the view that "the aimless play of the atoms, governed by the inexorable laws of mechanical causality, produced all phenomena in the world, inanimate, living, and mental." Polanyi [249, pp. 6-9] traces the origins of mechanism to Galileo's synthesis of a Pythagorean belief that the book of nature is written in geometrical characters, and Democritus' principle: "By convention coloured, by convention sweet, by convention bitter; in reality only atoms and the void." Even though Newton did not personally emphasise this in his philosophy, and many classical mechanists were troubled by the implications of action at a distance, Newton's followers unequivocally interpreted Universal Gravitation as the mechanistic ideal. Laplace, who made important extensions to Newton's physics, provided one of the more famous articulations of mechanism. It became known as Laplace's demon, after the following passage of *Essai philosophique sur les probabilitès* [198].

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

For all of its achievements, the Cartesian analytic method in conjunction with mechanism cast a shadow over subsequent Modern science, by setting the laws of theoretical physics as an ideal to which the less exact sciences should aspire, and to which they might eventually be reduced. In this view, only the mechanical properties of things were primary, and the properties studied in other sciences were derivative or secondary [249].

However, there were many areas of science that resisted the mechanical worldview. The debate between the mechanists and the vitalists on the distinction between inanimate matter and life dominated biology for three centuries. Roughly, the mechanists claimed that life was nothing-but chemistry and physics. Meanwhile, the vitalists countered that life was irreducible to the laws of physics, and the additional vital aspect of living organisms was capable of violating the laws of physics, or at least under-constrained by physics: life was in part self-determining. This was resolved only in the 1920s when the emergentists [6, 7, 62, 229–231] advanced a middle path that acknowledged the need for adherence to the laws of physics, while at the same time denying that all phenomena could be eliminatively reduced to physics. In retrospect, emergentism marked a pivotal advance towards systems thinking.

Checkland [81] identified three further classes of problems that persistently failed to be conquered by the Cartesian analytic method of science. They are problems of complexity, problems of social science, and problems of management. According to Checkland, the complexity that arises when densely interacting variables give rise to 'emergent phenomena' poses a serious problem, one that to date reductionist thinking and the experimental method has not been able to overcome.

Secondly, the problems of social science are not just densely interacting, but they contain individual human beings with potentially unique private knowledge and interpretations that condition their response, limiting the precision and general applicability of 'laws' governing social behaviour. In addition, social phenomena can always be interpreted from many more possible perspectives than are required for natural science.

Thirdly, Checkland cites management – the process of taking decisions in social systems – as problematic for analytic science. Operations Research (OR), the scientific approach to management, arose in support of military operational decisions during the second world war, and was institutionalised and applied widely in industry after the war. However, Checkland claims that OR's ability to solve problems with particular general forms has been of little use to real life problems, where the details that make a problem unique predominates any insight that knowledge of the general form provides. Even though Checkland does not substantiate this assertion, it is reinforced by observations from within the OR community that OR is being increasingly relegated to the tactical rather than strategic decision-making arena [117]:

> the uniqueness of the OR approach is not seen as indispensable, its methodology is challenged, it is regarded as a narrow specialist discipline, a suitable sanctuary for mathematicians, its involvement in implementation is tenuous and its general impact somewhat limited.

Whilst it would be a gross oversimplification to say that science prior to the mid twentieth century was non-systemic, it was not until this period that a self-consciousness of what it meant to be systemic arose. It is in the context of a dominant analytic approach to science, and a growing awareness of its limitations, that the systems movement sought to provide an alternative approach. Kline's Definition 2.3 of **system** showed that the term already had a common usage within analytic science as the object or objects of interest. However, this usage does not attribute any properties to a system – it acts as a container, merely a convenient label for the collection of interacting objects under investigation. In contrast, the systems movement began by articulating a more substantial conception of system.

3.3 Enter the System

In the West, von Bertalanffy's [308] seminal paper on open systems, published in 1950, is usually attributed as seeding the rise of the systems movement, although he had published on a systems approach to biology since 1928 [307]. The purpose of the paper was to rigorously account for the key distinction between the organismic systems of biology, compared with the closed systems of conventional physics. In making this distinction, von Bertalanffy wished to scientifically account for apparent paradoxes in the characteristics of living systems, when considered in relation to the physics of closed systems. For example, the second law of thermodynamics states the inevitability of increasing entropy and the loss of order, and yet the evolution and development of biological systems can exhibit sustained increases in order. By providing empirically supported scientific explanations of phenomena such as equifinality (when a system's dynamics are not state-determined) and anamorphosis (the tendency towards increasing complication), von Bertalanffy simultaneously extinguished vitalism and firmly established the roots of systems theory in the natural sciences.

Von Bertalanffy's emphasis on flows of matter (and later energy and information) into and out of an open system brought attention to the environment of the open system. This adds something that Definition 3.1 does not contain: a system is more than just a set of components and their relationships – it is a complex whole that affects and is affected by its environment. Further, a system has a boundary that prevents it from becoming mixed with its environment. The implication of the environment is that a system must always be understood in context.

The domain of systems theory does not cover all systems – it was never intended as a theory of everything. In 1948, Weaver [323] noted that between mechanics and statistical mechanics, there was an absence of mathematical techniques for the treatment of systems with medium numbers of components, whose behaviour could be much more varied than either simple machines or very large ensembles. Weaver gave a name to the systems in this gap, saying that "one can refer to this group of problems as those of *organized complexity*". Weaver illustrated the domain of organised complexity graphically. A slightly elaborated version of this graphic by Weinberg [324] is reproduced in Figure 3.1. In Weaver's view, the entities that systems theory studies are substantial (physical) systems. Occasionally, confusion has arisen as to whether the domain of systems theory extends to the study of abstract (conceptual) systems. (Ashby [20, p. 260] notes this general uncertainty within GST.) This confusion is exacerbated because systems theories of both persuasions have been developed, often without explicitly addressing their domain of applicability. However, it is fair to say that both substantial and abstract systems

NOTE: This figure is included on page 54 of the print copy of the thesis held in the University of Adelaide Library.

Figure 3.1: Types of systems, after Weinberg [324].

have been legitimate subjects of systems enquiry.

Checkland [81, pp. 74-92] suggested that systems thinking was founded on two pairs of ideas:

1. Emergence and hierarchy, originating in organismic biology and generalised in GST; and

2. Communication and control, originating in communication engineering and generalised in cybernetics.

To these pairs of ideas should be added the requirement for a systems approach to be broader than any conventional discipline: a systems approach is fundamentally interdisciplinary. In retrospect, a sophisticated theory of systems that preceded both GST and cybernetics had been developed by Bogdanov [114] in Russia and published sometime between 1910-1913, but for political reasons his Tektology was not widely known until after the Cold War, by which time the systems movement was already well established in the West, and synonymous with GST and cybernetics.

Another precursor to a recognised systems movement (by about a decade) was Angyal's [12] theory of systems, developed in the context of psychology. In particular, Angyal distinguished between relations and systems:

A relation requires two and only two members (relata) between which the relation is established. A complex relation can always be analysed into pairs of relata, while the system cannot be thus analysed.

This distinction leads to the difference between aggregates, in which the parts are added, and systems, in which the parts are arranged. Note that Newtonian mechanics only describes the behaviour of aggregates. Angyal also realised the importance of the manifold that the system is embedded in, saying that a system is "a distribution of the members in a dimensional domain." Hence, a system is more than a set of interrelated components: the relations must be *organised*. Spatial arrangements are an important determinant of systemic properties. This also implies that many combinations of relations will not be possible for any particular system, since the relations must conform to the system's organisation. Thus organisation imposes order on a system, which can be thought of as a constraint on its dynamics.

Because the systems approach stood in contrast to mechanism (and also the related 'isms' of atomism and individualism), some began to associate systems with holism. The systems philosopher Bunge [69, 71] recognised this conflation and sought to distinguish systemism from holism. According to Bunge [69],

> a holistic philosophy is one that not only accepts the thesis that a whole is more than a mere aggregation of its parts: holism maintains also that wholes must be taken at face value, understood by themselves, not through analysis.

Because the holistic approach rejects the possibility of analysis, it relies upon the method of intuition, not rational explanation or empirical experiment. While the systems approach recognises the existence of emergent properties, it nevertheless seeks to explain them in terms of how their constituent parts are organised. Where holism is satisfied with a non-rational apprehension of un-analysed wholes, systemism aims to demystify emergent properties by providing scientific understanding that utilises analysis as well as synthesis. Therefore, it is equally important that the systems approach be distinguished from holism as from mechanism.

Another important refinement to the philosophical characterisation of the systems approach was provided by Churchman [83], in the form of the following question:

> How can we design improvement in large systems without understanding the whole system, and if the answer is that we cannot, how is it possible to understand the whole system?

Ulrich [302], a student and self-proclaimed disciple of Churchman, realised that Churchman's question captured "the real challenge posed by the systems idea: its message is not that in order to be rational we need to be omniscient but, rather, that we must learn to deal critically with the fact that we never are." Thus, in a very deep sense, the systems approach is tied to understanding the limits of representations.

3.3.1 General systems theory

Von Bertalanffy together with Rapoport and Boulding formed the Society for General System Theory in 1954, which organised the systems community and provided a yearbook from 1956 dedicated to systems research. Along with the closely related field of cybernetics, GST helped to define the core principles of the systems approach.

For von Bertalanffy, the main propositions of GST (adapted from [309]) were:

1. Isomorphisms between the mathematical structures in different scientific disciplines could integrate and unify the sciences;
2. Open systems require consideration of the flow of energy, matter and information between the system and its environment;
3. Within open systems, the same final system state may be reached from different initial conditions and by different paths – open systems exhibit equifinality;
4. Teleological behaviour directed towards a final state or goal is a legitimate phenomenon for systemic scientific inquiry;
5. A scientific theory of organisation is required, to account for wholeness, growth, differentiation, hierarchical order, dominance, control and competition; and
6. GST could provide a basis for the education of scientific generalists.

Von Bertalanffy's principle concern was to provide an alternative foundation for unifying science, which he proposed in reaction to the reductionist mechanistic worldview. In particular, he rejected the crude additive machine theories of organic behaviour, which treated wholes as nothing more than linear aggregates of their components. It is notable that von Bertalanffy [309, 310] really only mentions emergence and hierarchy in passing. GST adopted almost unchanged the theory of biological emergentism developed in the 1920s, while the task of developing hierarchy within GST was taken up by other authors. In fact, I would suggest that rather than a theory of general systems, GST resulted in the more modest contribution of several theories of hierarchies. The remainder of this section on GST will consider the twin concepts of emergence and hierarchy.

Emergence as "the whole is more than the sum of the parts" justified the need to understand systems in addition to understanding their parts. By understanding emergent properties, the general systems theorists felt that they could offer insights that the reductionist / mechanist agenda ignored, because mechanism could not account for the non-additive relationships between components. The ability to explain emergent properties is a prerequisite for a general theory of organisation. A detailed history of emergence is deferred to the chapter on emergence, where it is given in Section 6.2.

The hierarchical nature of systems is a logical consequence of the way system is defined in terms of its constituent parts, since the parts may also meet the definition of system. This was noted by many authors, but Boulding and Simon made particularly influential extensions to this corollary. Boulding's [55] *Skeleton of Science* presented a hierarchical view of systems, "roughly corresponding to the complexity of the 'individuals' of the various empirical fields." Where von Bertalanffy [308] had focussed on the relationship of physics and chemistry with biology, Boulding generalised this into a hierarchy with nine levels, which began with physics and extended through biology, psychology, sociology and metaphysics. Following Matthews [216, p. 201], Boulding's hierarchy is summarised in tabular form in Table 3.1.

Boulding saw these levels as different layers of theoretical discourse. That is, the layers were not part of nature itself, but were a way of organising the theoretical concepts and empirical data collected in the various sciences (although it has not always been presented this way by other authors). When Boulding later aligned specific disciplines with the levels, the intent was to highlight that, for example, any analysis of level 8 sociocultural systems using only the methods of level 1 and 2 structures and mechanics would be necessarily incomplete. Boulding's hierarchy has been extended by Checkland [81] and critiqued by numerous authors including Capra [73]. Despite its shortcomings, Boulding's hierarchy was an influential representation of the relations between the sciences. It also reveals a view that many early systems theorists shared: that their role was to be one step removed from empirical science, in order to recognise the broader patterns occurring across science. Others, such as Bunge [67] and Rapoport [258], later developed more explicit arguments that the systems approach lies midway between the scientific and philosophical approaches.

Table 3.1: Boulding's hierarchy of systems complexity.

Level	**Characteristics**	**Examples**	**Relevant Discipline**
1. Structure	Static	Crystals	Any
2. Clock-works	Pre-determined motion	Machines, the solar system	Physics, Chemistry
3. Control mechanisms	Closed-loop control	Thermostats, mechanisms in organisms	Cybernetics, Control Theory
4. Open systems	Structurally self maintaining	Flames, biological cells	Information Theory, Biology (metabolism)
5. Lower organisms	Organised whole with functional parts, growth, reproduction	Plants	Botany
6. Animals	A brain to guide total behaviour, ability to learn	Birds and Beasts	Zoology
7. Humans	Self-consciousness, knowledge symbolic language	Humans	Psychology, Human Biology
8. Socio-cultural systems	Roles, communication, transmission of values	Families, clubs, organisations, nations	Sociology, Anthropology
9. Transcendental systems	Inescapable unknowables	God	Metaphysics, Theology

Simon's [292] classic paper on *The Architecture of Complexity* sought to explain some features of naturally arising hierarchies, in contrast to Boulding's attempt to relate the description of nature within different disciplines of science. Simon proposed that systems with a large number of non-simply interacting parts frequently take the form of hierarchy, defined as

> a system that is composed of interrelated sub-systems, each of the latter being, in turn, hierarchic in structure until we reach some lowest level of elementary subsystem.

Simon is describing a nested hierarchy, which was later generalised in the highly abstract GST dialect of hierarchy theory [9] that this paper helped to initiate. There are two central insights in [292]. Firstly, what Simon saw as the ubiquity of hierarchies in natural complex systems is explained by a simple probability argument: the time it takes for the evolution of a complex form depends critically on the number of potential intermediate stable forms, as these building blocks to a large degree insulate the process of system assembly against the effects of environmental interference. Given that a hierarchy of building blocks can be assembled orders of magnitude faster than a non-hierarchical assembly process, among complex forms, hierarchies are the ones that have the time to evolve.

Secondly, Simon realised that the interactions at each level of the hierarchy are often of different orders of magnitude, and commonly the interactions are strongest and most frequent at the lowest level in the hierarchy. When these conditions hold, the hierarchy is *nearly decomposable*, which simplifies the analysis of a complex system in several ways. Near decomposability implies that subparts belonging to different parts only interact in aggregate, meaning individual interactions can be ignored: the levels are screened off from each other by rate differences. This is why in modern hierarchy theory, three levels of a hierarchy, one up and one down, are generally considered sufficient for analysis of the focal level. Also, nearly decomposable hierarchies are highly compressible: they have significant redundancy, meaning that a relatively small amount of information may adequately describe an apparently complex system.

3.3.2 Cybernetics

At about the same time that GST was attempting a general mathematics of organisation, the new field of cybernetics was embarking on a similar quest to uncover the general mathematics of machines. According to Wiener [326], who suggested the name for the field,

> cybernetics attempts to find the common elements in the functioning of automatic machines and of the human nervous system, and to develop a theory which will cover the entire field of control and communication in machines and in living organisms.

In [326], Wiener described the initial results from the cybernetic community, whose impressive interdisciplinary role-call included Rosenblueth, von Neumann, McCulloch, Pitts, Lorente de Nó, Lewin, Bateson, Mead and Morgenstern. With the exception of Morgenstern, they were all members of the core group of ten conferences on cybernetics held between 1946 and 1953, sponsored by the Josiah Macy, Jr. Foundation and chaired by McCulloch. Wiener identified the principle of feedback as a central step towards the study of the nervous system as an integrated whole, outlining the role of negative feedback in control, and positive feedback in producing instability. The principle of feedback distinguishes cybernetic systems by their circular organisation, where activity flows from effectors acting on the environment, to sensors detecting their outcome, which then act upon the effectors, thereby closing the feedback loop. The other major distinction that Wiener drew was between communication engineering and power engineering as a basis for understanding mechanical systems. He identified a change in emphasis in the former from the economy of energy to the accurate reproduction of a signal. The implications of this distinction are that understanding the behaviour of a control system depends not on the flow of energy so much as the flow of *information.* Fortuitously, at the same time Shannon [287] was developing a quantitative definition of information as the average reduction in uncertainty (see Chapter 5), which became one of the dominant frameworks within cybernetics, and more generally within systems theory.

Although Wiener's [326] analogies between calculating devices and the human brain now appear dated, a number of the big systems ideas arose from Wiener's cybernetic collaborators in attempting to understand the behaviour of "all possible machines". They include von Neumann's self-reproducing automata [316], McCulloch and Pitts' [219] model of the neuron that forms the basis for artificial neural networks, Maturana's [217] autopoesis, von Foerster's [314] second order cybernetics and Bateson's [36] ecology of mind, each exploring different implications of circularity.

Outside the US, the British psychiatrist Ashby had published on what became known as cybernetics from 1940, inventing the homeostat – a machine that adapted towards self-regulation – and writing the 1956 classic text *An Introduction to Cybernetics.* Ashby's ideas have had a profound influence on complex systems theory, pre-empting and shaping the complex systems approach, and still providing insights on contemporary concerns such as self-organisation, adaptation and control. Perhaps Ashby's greatest contributions were the connections he developed between information theory and systems theory. In [18, p. 3], Ashby observed

> cybernetics typically treats any given, particular, machine by asking not "what individual act will it produce here and now?" but "what are *all* the possible behaviours that it can produce?"
>
> It is in this way that information theory comes to play an essential part in the subject; for information theory is characterised essentially by its dealing always with a *set* of possibilities ...

Ashby's most famous result, introduced in this book, applies information theory to machine regulation to yield the law[6] of requisite variety. The law provides a quantitative measure of the amount of variety a regulator R must absorb in order to produce goal-directed behaviour. The story of representation as the problem the brain is intended to solve, introduced in Section 2.5, is augmented by Ashby's

[6]The law of requisite variety is actually a mathematical theorem, not a physical law.

understanding of the *purpose* of constructing representations that shape behaviour. That purpose is primarily to maintain essential system variables, E, within certain limits. Ashby argues that natural selection has eliminated those organisms that are not able to regulate the flow of environmental variability. The remaining species of organisms all employ mechanisms that actively resist environmental disturbances, D, that would push the essential variables outside their limits and result in death.

The law is simple to introduce. If different disturbances can be countered by the same response by the regulator, they are not counted[7] as distinct members of D. If the set of possible outcomes, O, of the disturbance followed by regulation (denoted by the transformation matrix T) is within the acceptable limits of E, it is compatible with the organism's continued existence (denoted by the set η). In this case, the regulator is considered successful. Let the variety, measured as a logarithm of the number of possibilities, of D, R and O, be V_D, V_R and V_O respectively. Then the law of requisite variety is simply

$$V_O \leq V_D - V_R. \tag{3.2}$$

Intuitively, goal-directed behaviour demands that variety in the outcome remain below some bound. In order to reduce V_O, V_R must be increased: only variety in the regulator can destroy variety created by the environmental disturbance. Another[8] way to understand the law of requisite variety is to ask what if $R > D$?. Then by Equation 3.2, $V_O < 0 \Rightarrow O < 1$. This means that the disturbances are perfectly countered by the regulator, and so there is no variety in the outcome of the process.

The law of requisite variety sheds light on the nature of control. Figure 3.2, adapted from [18], shows the causal influences between the variables introduced above and a new control variable C. In this model, C decides the desired outcome or sequence of outcomes, which R must obey. Roughly, C, R and E constitute the organism or the 'system', which is open to energy and information but closed with respect to control. R takes information from the independent sources, C and D, and attempts to produce an action that achieves C's objective in spite of the external disruption D. The 'environment' T transforms the actions of D and R, which influences the value of the essential variables E. Therefore, control over E *necessarily requires* regulation. This can be interpreted as the communication of a message from C to E via the compound channel T, while transmitting nothing from D. This is not just a superficial analogy to information theory: the law of requisite variety is a restatement of Shannon's [287] tenth theorem on channel coding. This reveals the deep connections between goal-directed behaviour (see Section 2.6), control, regulation and communication.

Another connection between information theory and systems is evident in Ashby's definition of complexity as the quantity of (Shannon) information required to describe a system [21], although Ashby was not the first to propose this definition.

Where GST predominantly analysed structure to understand organisation, cybernetics developed a complementary approach for analysing dynamical *behaviour*,

[7]This way of counting disturbances does a lot of the work in the law of requisite variety, and it is the key assumption that must be satisfied for real world applications of Ashby's law to be valid.

[8]Thanks to Scott Page for this suggestion.

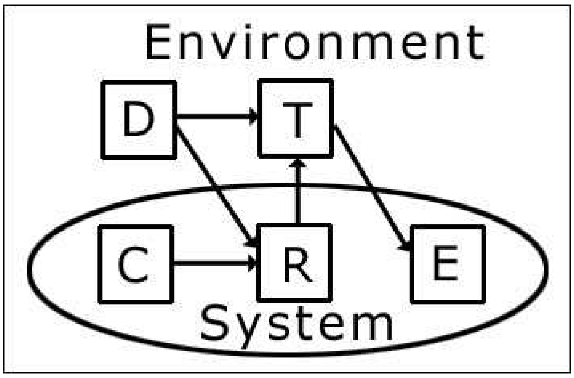

Figure 3.2: Control and regulation in Ashby's law of requisite variety.

independently of how the system was internally structured. The two are complementary because structure can be viewed as a constraint on the system's dynamics, while the dynamics are responsible for the formation of structure. The behaviourist approach to modelling systems was formalised within cybernetics by generalising the electrical engineer's 'black box'. The black box approach assumes that the internal structure of a system is hidden, and so knowledge of the system must be obtained by systematically varying inputs to the system and observing the corresponding outputs. The same approach to behavioural psychology is described as stimulus-response (as in [321]), and both terminologies were used interchangeably in cybernetics.

Although GST and cybernetics were strong allies in advancing the systems movement, there were some important differentiators. The research agenda for cybernetics was focussed on machines, while GST was much broader. The commitment of cybernetics to behaviourist mechanical explanations outlined in the previous paragraph stands in contrast to the position within GST that mechanics provides an incomplete account of open systems. Thus in Boulding's hierarchy, systems at or above level four require more than just an understanding of information and control as provided by cybernetics at level three. Also, in cybernetics emergence was typically dismissed as incompleteness of the observer's knowledge of the parts and their couplings (see for example [18, pp. 110-112]). The only time cybernetics acknowledged genuine emergence was as a result of self-reference or circularity. Unlike GST, cybernetics embraced mechanistic explanations, but sought to augment the physicists' view of causes preceding effects with circular causation and goal-directed behaviour.

3.3.3 Systems analysis

Systems analysis was an extension of Operations Research (OR) to broader concerns. Morse and Kimball's [232] book set the standard for OR, including its definition.

> Operations research is a scientific method of providing executive departments with a quantitative basis for decisions regarding the operations under their control.

Although slightly older, OR had some common ground with the new systems approaches. In particular, OR too saw itself as an interdisciplinary approach based on scientific methods. The way in which OR provided a scientific basis for decisions was to use mathematical models. Also, OR often developed integrated solutions, rather than considering individual components in isolation. Network flow, queueing theory and transportation problems are typical examples of this, but also serve to highlight some key differences. Traditional OR addressed problems where the objectives were precisely given, and the components were fixed [128], but their configuration could change according to the value of the control parameter $\mathbf{x}$. The operations researcher would structure the problem by framing it in such a way that the performance of alternative configurations could be compared against the same objective. This focus on increasing measurable efficiency meant that OR was traditionally interested in only one organisation: the configuration that gives the global minimiser of the objective function $f(\mathbf{x})$. As OR developed increasingly sophisticated mathematical techniques, it became more theoretical and less interdisciplinary, which led to a number of issues that have already been mentioned in Section 3.2.

In the newly established RAND (Research ANd Development) Corporation, the mathematician Paxson (whose work has not been declassified) was more interested in decisions affecting the next generation of military equipment, than configuring a fixed set of platforms already in operational service. Early applications of OR to what Paxson called "systems analysis" were criticised because they did not adequately consider costs [110]. Consequently, systems analysis became a collaborative venture between mathematicians/engineers and economists. Successful high profile projects in air defence and the basing of bombers added "Red Team" game-theoretical approaches to the systems analysis methodology, generated sustained interest in systems analysis within RAND, and gained the attention of senior management. As well as borrowing from and extending OR, systems analysis inherited the black boxes and feedback loops from cybernetics in order to construct block diagrams and flow graphs. A subsequent technique for performing systems analysis – Forrester's [127] system dynamics[9] – made more explicit use of these cybernetic concepts. After a decade where RAND led the formalisation and application of the systems analysis methodology, in 1961 the Kennedy Administration brought both RAND people and systems analysis techniques into government to provide a quantitative basis for broad decision-making problems [110]. McNamara's use of systems analysis within the Department of Defense and NATO was highly controversial and widely criticised. Nevertheless, it is notable as the systems approach which has had the greatest impact on society, due to the scale and nature of the decisions that it justified. In particular, the Vietnam war was planned with considerable input from systems analysis techniques combining operations research with cost analysis.

Aside from a rudimentary use of feedback, systems analysis was largely independent of developments in systems theory. As Checkland [81, p. 95] nicely described it,

> on the whole the RAND/OR/management science world has been un-

[9]System dynamics represented the world as a system of stocks and flows, from which the behaviour of feedback loops could be deduced.

> affected by the theoretical development of systems thinking, it has been *systematic* rather than *systemic* in outlook ...

By this, Checkland meant that systems analysts assembled cookbooks of mechanical solutions to classes of recurring problems, rather than developing techniques that addressed emergent properties, interdependencies and environmental influences. In so far as it was applied to the class of simple mechanical systems depicted above in Figure 3.1, this was perfectly reasonable. However, as the interests of systems analysis broadened from merely technical systems to include many social factors, this omission became more apparent.

3.3.4 Systems engineering

Systems engineering has a history quite separate to the systems movement. Its closest historical link comes from the application of systems analysis techniques to parts of the systems engineering process. The need for systems engineering arose from problems in the design and implementation of solutions to large scale engineering challenges spanning multiple engineering disciplines. A multidisciplinary team of engineers required a lead engineer whose focus was not the design of individual components, but how they integrated. Consequently, management concerns were as significant as technical challenges for a systems engineer. The Bell Telephone Laboratories and Western Electric Company's design and manufacture of the Nike air defence system, commenced in 1945, is widely cited as one of the first systems engineering projects. The surface to air missile defense program integrated ground-based tracking radars, computers and radio controlled anti-aircraft rockets, in order to protect large areas from high altitude bombers. It was novel because unlike conventional anti-aircraft artillery, Nike allowed continuous missile guidance: the radars and computers enabled feedback and control. Bell Labs were the prime contractor for the project, while much of the detailed engineering was undertaken by the major subcontractor, Douglas Aircraft Company. The 1945 Bell Labs report *A Study of an Antiaircraft Guided Missile System* was considered a classic in applied interdisciplinary research due to its depth of insight, scope, and influential role in the systems engineering of Nike [120].

Following the success of individual systems engineering projects, Bell Labs structured itself around the new systems engineering approach. Bell Labs was organised into three areas: basic research, systems engineering and manufacturing projects [178]. The systems engineering area provided the interface between advances in communications theories and the manufacture of commercial systems. Because of the "whole system" perspective within the systems engineering area, it was responsible for prioritising the activation of projects with the greatest potential user benefit, within the technical feasibility of communications theory. The responsibility of the systems engineer was the choice of the "technical path" between theory and application in order to create new services; improve the quality of existing services; or lower their cost. Because of its emphasis on user benefit, standards were seen to play a vital role in systems engineering. Standards were used to measure quality of service, which enabled cost benefit analysis of different technical paths.

When framed as the decision-making problem of selecting between different technical paths on the basis of cost effectiveness, systems engineering appears closely related to the problems that RAND were concurrently tackling with systems analysis. Bell Labs was aware of this analogy [320], and drew on OR and systems analysis techniques. However, the systems engineer typically placed a greater emphasis on technical knowledge than the systems analyst, and correspondingly less emphasis on mathematical models. In contrast to the dynamical models of systems analysis, systems engineers developed architectures to represent system designs.

Outside Bell Labs, Project Apollo was one of the highest profile early successes of the systems engineering approach, which quickly spread from its origins in defence to also become the standard approach to large scale civilian projects. The traditional systems engineering process can be summarised as follows: 1) Customer needs are captured in precise, quantified requirements specifications; 2) System requirements are decomposed into requirements for subsystems, until each subsystem requirements is sufficiently simple; 3) Design synthesis integrates subsystems; and 4) Test and evaluation identifies unintended interactions between subsystems, which may generate additional requirements for some subsystems. If there are unintended consequences (i.e. unplanned emergent properties), the process returns to stage 2, and repeats until the system meets the requirements.

Because systems engineering is applied to the most ambitious engineering projects, it also has a long list of large and public failures [29]. Partly in response to perceived failures, systems engineering has identified "System of Systems" problems as a distinct class of problems that are not well suited to traditional centrally managed systems engineering processes. Although there is not a standard System of Systems (SoS) definition, the term SoS usually denotes heterogeneous networks of systems including a majority of Maier's [215] discriminating factors: operational and managerial independence, geographical distribution, and emergent and evolutionary behaviours. The recent popularity of the SoS buzzword in the systems engineering literature has prompted the expansion of systems engineering techniques to include methods that can cope with evolving networks of semi-autonomous systems. This has led many systems engineers to read more widely across the systems literature, and is providing a re-conceptualisation of systems engineering as part of the systems movement, despite its historical independence. This is reflected in the latest INCOSE handbook [148, p. 2.1], which states "[t]he systems engineering perspective is based on systems thinking", which "recognizes circular causation, where a variable is both the cause and the effect of another and recognizes the primacy of interrelationships and non-linear and organic thinking—a way of thinking where the primacy of the whole is acknowledged".

3.3.5 Soft systems

The systems approaches I have surveyed so far have each expressed an unbounded enthusiasm for multi-disciplinarity in their missionary papers, and yet all have tended to converge towards a relatively narrow band of 'hard' scientific methods, where this is taken to mean strong scepticism towards any theory that cannot be stated exactly in the language of mathematics. C. P. Snow's famous "Culture Gap" between literary intellectuals and scientists proclaimed that a gulf between the two cultures prohibited any real communication or shared understanding between them. A related, but perhaps deeper cultural divide was noted a generation earlier by James [170] (as quoted in [216]) in 1909:

> If you know whether a man is a decided monist or a decided pluralist, you perhaps know more about the rest of his opinions than if you give him any other name ending in 'ist' . . . to believe in the one or in the many, that is the classification with the maximum number of consequences.

Even within science, this metaphysical divide can be observed. Another way of approaching the dichotomy is to ask whether the goal of enquiry is objective knowledge or inter-subjective discourse. Hard science, by accepting only mathematically precise arguments, adopts a monistic stance. Consequently, hard science is associated with the search for a *unified rational foundation* for systems universally valid across the sciences. Von Bertalanffy, Wiener, Ashby and Forrester all clearly held this aspiration to varying degrees. In contrast, a pluralist position rejects the possibility of unification, emphasising the incommensurability of frameworks that view reality from different perspectives. This is rejected by the monist, because it is seen to be opening the door to inconsistency, paradox, relativism and irrationality. Within the Macy conferences, issues of subjective experience and *Gestalten* (form perception) were raised, usually by psychoanalysts such as Kubie [10]. The hard scientists sustained vocal criticism that tended to suppress these ideas: in particular, they strongly questioned the scientific status of psychoanalysis.

Von Foerster, an editor of the Macy conference proceedings, later developed second order cybernetics [314], which shifted attention from the cybernetics of *observed* systems, to the cybernetics of *observing* systems. In studying the "cybernetics of cybernetics", von Foerster applied cybernetic concepts to the observer, which led him to reject the possibility of objective observation. Von Foerster's [315] variant on James' metaphysical distinction was introduced as

> a choice between a pair of in principle undecidable questions which are, "Am I *apart from* the universe?" Meaning whenever I *look*, I'm looking as if through a peephole upon an unfolding universe; or, "Am I *part of* the universe?" Meaning whenever I *act*, I'm changing myself and the universe as well.

By adopting the latter position, von Foerster was dismissing the search for a unified rational foundation for systems as an ill-posed problem. Even though von Foerster approached second order cybernetics using essentially hard techniques, he played an important role in undermining the doctrine of objectivity that "[t]he properties of the observer shall not enter the description of his observations" [314]. Second order cybernetics provides a scientific basis for the theory-ladenness of observation: the philosophical assertion that all observations contain an inseparable element of theory.

In 1981, Checkland [81] published the first sophisticated systems methodology for 'soft' problems. Checkland firstly divided systems into five types: natural systems (atomic nuclei to galaxies); designed physical systems (hammers to space rockets); designed abstract systems (mathematics, poetry and philosophy); human activity systems (a man wielding a hammer to the international political system); and transcendental systems (as in Boulding's hierarchy, although these are not considered in detail). Checkland's central thesis was that human activity systems were fundamentally different in kind from natural systems, and consequently required a fundamentally different approach. Checkland [81, p. 115] argued that

> The difference lies in the fact that such systems could be very different from how they are, whereas natural systems, without human intervention, could not. And the origin of this difference is the special characteristics which distinguish the human being from other natural systems.

That special characteristic is self-consciousness, which is claimed by Checkland to provide "irreducible freedom of action" to humans. However, one should note that this distinction between the human being and other natural systems is much more of a metaphysical assertion than an empirical fact, and it is the most crucial theoretical assertion for Checkland's argument against the use of hard methods in human activity systems.

The fundamentally different approach that Checkland used to investigate human activity systems was 'action research'. According to Checkland [81, p. 152],

> Its core is the idea that the researcher does not remain an observer outside the subject of investigation but becomes a participant in the relevant human group.

From this, it is clear that action research follows von Foerster's lead in exploring the implications of *acting* as *part of* the universe. This commitment is responsible for the two distinguishing features of the soft systems approach: the way problems are framed, and the way models of the system are built and used.

The difference in problem framing is most clearly visible in Checkland's [81, p. 316] definition of a soft problem:

Definition 3.3 (Problem, Soft)**.** *A problem, usually a real-world problem, which cannot be formulated as a search for an efficient means of achieving a defined end; a problem in which ends, goals, purposes are themselves problematic.*

Thus, instead of defining an objective that can be tested at any point in time to see if it has been met, the soft systems approach operates with a comparatively unstructured and fuzzy understanding of the "problem situation". Because the problem is framed using only minimal structure, the quantitative methods of OR and systems analysis cannot be brought to bear on its solution. Nor is the aim of soft systems optimal efficiency, instead it seeks feasible, desirable change.

The way models are built is fundamentally pluralist: "human activity systems can never be described (or 'modelled') in a single account which will be either generally acceptable or sufficient" [81, p. 191]. It is accepted that different people will view the system and its problems differently, and models aim to make explicit the assumptions of a particular view in order to facilitate dialog, rather than building a single foundational representation of the system. Thus, the difference between hard and soft systems approaches is a deep divide: it is James' metaphysical distinction between the belief in the one or in the many. Perhaps the most important point in favour of soft systems methodology is that it has mostly been considered useful in practice for solving real-world problems [225].

Habermas [141] added a further category in contrast to what I have described as hard and soft approaches, with the development of critical theory. According to Habermas, humans have developed a technical interest in the control and manipulation of the world; a practical interest in communicating to share understanding with other people; and an emancipatory interest in self development and freedom from false ideas. Critical theory aims to reveal systematic distortions resulting from the power structure that affect both hard (technical) and soft (communicative) approaches. It formed the theoretical basis for critical systems theory, which has produced several methodologies [168, 303] for taking practical action in management contexts. However, because the intention of critical theory is emancipation, not design or even intervention, these methods have struggled to resolve this fundamental inconsistency. The main contribution of critical systems theory to date has been to critique both the hard and soft systems approaches, rather than to advance a practical alternative to systems design (see for example [224]). Consequently, it will not be considered further in this thesis.

3.3.6 Complex systems

Complex systems became known as a distinct systems discourse with the establishment of the Santa Fe Institute (SFI). A widely read romantic account of the formation of the SFI and its initial research agenda was published by the journalist Waldrop [319]. Its success set the model for communicating complex systems research to the informed general public through easy to read popular science novels. These novels attempted to capture the excitement of research on the frontier of "a new kind of science", while at the same time emphasising its applicability to important global issues, which are inevitably both systemic and complex. The upside of the way complex systems was marketed is that it has grown to become the most active[10] area of systems research; the barrier to entry from many disciplinary backgrounds is comparatively low, since discipline-specific jargon is minimised; and there is a largely positive perception of complex systems in the wider community, where there is awareness at all. The downside to this approach includes a proliferation of poorly defined buzzwords; a level of hype that is reminiscent of early AI; and widespread use of complex systems as a superficial metaphor. Of course this is not the only legacy of the SFI: it has established a dynamic network of talented systems researchers affiliated with many Universities, which interacts with a network of powerful corporations (both networks are almost entirely confined to the USA), within a unique and effective business model. It has also published a high quality series of proceedings from invite-only workshops on a diverse array of interdisciplinary topics that have added considerable substance to the complex systems approach.

The New England Complex Systems Institute (NECSI) has taken a much more systematic approach to complex systems. Bar-Yam's [25] complex systems textbook

[10]This claim is justified by the scale and number of conferences, the number of complex systems centres, and the volume of new research, in comparison to other systems approaches.

collated a large survey of the mathematical techniques in complex systems, including iterative maps, statistical mechanics, cellular automata, computer simulation, information theory, theory of computation, fractals, scaling and renormalisation. These techniques were then further developed and applied to the brain, protein folding, evolution, development, and human civilisation, to demonstrate the interdisciplinary breadth of complex systems applications. In addition to publishing the only comprehensive textbook on complex systems[11], NECSI hosts the International Conference on Complex Systems, which is the premier complex systems conference.

Taking these two organisations as representative of the complex systems approach, complex systems is essentially a refinement of the GST/cybernetics research agenda. There has, in fact, been remarkably little interaction between complex systems and second order cybernetics, soft systems or critical systems: the insights from these alternative systems approaches are neither acknowledged nor addressed. (GST is usually just not acknowledged – the term never appears in [25, 132, 175, 197, 204, 281, 319].) Nor do the contemporary soft approaches draw on the techniques of complex systems. This could be explained by a general aversion to mathematics – for example, Midgley's four volumes of systems thinking explicitly aimed to exclude "papers that were heavily dependent on a high level of mathematical reasoning" [222, p. xxi]. Consequently, of 76 papers on systems thinking, only one high-level paper on complex systems was included. There has been greater interaction between complex systems and other 'hard' systems approaches: OR and systems analysis share techniques such as genetic algorithms [135] and particle swarm optimisation [179] with complex systems, while the new area of complex systems engineering [59] augments systems engineering problems with complex systems approaches. To show just how closely the aims of complex systems overlap with the GST/cybernetics research agenda, analogs of the ideas introduced in Sections 3.3.1 and 3.3.2 are now identified within complex systems.

Whereas von Bertalanffy imagined unity of the sciences through isomorphisms, Gell-Mann [132, p. 111] talks of building staircases (or bridges) between the levels of science, while Bar-Yam [25, p. 2] cites universal properties of complex systems, independent of composition. Self-organisation, which has been labelled antichaos [174] due to its stabilising dynamics, is a substitute for equifinality. However, equal attention is now given to chaotic dynamics – which exhibit sensitivity to initial conditions – to demonstrate the limits of the classical state-determined system. Goal-directed behaviour is studied under the heading of autonomous agents [176, p. 49]. Organisation is still of central importance [176, p. 81] – especially self-organisation [175, 280], [25, p. 691] – which largely follows in the tradition of either Prigogine [235] (viewed as an open system far from equilibrium) or Ashby [20] (viewed as an increase in organisation or an improvement in organisation).

The preoccupation in complex systems with quantifying the complexity of a system is a minor extension of the desire to understand organisation in general, and qualitatively expressed in Weaver's concept of organised complexity. The most common way to quantify complexity is by the amount of information it takes to describe a system, which follows from the use of Shannon information theory in cybernetics,

[11] Other more specialised books do now exist, such as Boccara's [51] textbook on the physics of complex systems.

but also borrows from theoretical computer science the notion of algorithmic complexity (see Section 5.3 for detailed examples of measuring complexity). Ashby's conceptualisation of adaptation and learning as an adaptive walk in the parameter space of a dynamical system is followed by Kauffman [175, p. 209], and generalised theories of adaptation and evolution have been developed by Holland [153] and Bar-Yam [25, p. 531] respectively. Meanwhile, genetic algorithms [135], artificial neural networks [159], reinforcement learning [298] and ant colony optimisation [112] are a selection of practical techniques used to generate adaptation in complex systems.

Von Neumann's [316] theory of natural and artificial automata are still studied in a mostly pure mathematical branch of complex systems – cellular automata – that can be expected to gain increasing attention from potential applications in areas such as Field Programmable Gate Array (FPGA) design and nanotechnology. Emergence continues to be a central concern in complex systems [30, 99, 154]. Boulding's hierarchy among the disciplines of science is maintained by Gell-Mann [132, pp. 107-120], although Simon's view of systems as nested hierarchies has been mostly supplanted by the consideration of systems as networks [35, 322]. An approximate translation between GST/cybernetics and complex systems is summarised in Table 3.2.

Table 3.2: Comparison of the themes of general systems theory and cybernetics with complex systems.

General Systems Theory and Cybernetics	**Complex Systems**
Unity of science through isomorphisms	Coherence of science through bridges
Isomorphic mappings	Universality classes
Emergence	Emergence
Organisation	Self-organisation
Organised complexity	Complexity
Adaptation	Adaptation/evolution
Equifinality	Chaos and antichaos
Goal-directed behaviour	Autonomous agents
Automata	Cellular automata
Hierarchies	Networks

With so much commonality, what is different about the field of complex systems? In terms of the research agenda, *very little*.

Phelan [247] notes the broad similarities between complexity theory (referred to here as complex systems) and systems theory (by which Phelan means systems methodologies in operations research, engineering and management science, following from the pioneering systems work of von Bertalanffy, Ashby and Boulding), but attempts to also distinguish between the two. According to Phelan, systems theory is predominantly focused on intervention, whereas complex systems is more interested in exploration and explanation. Whereas the system dynamics models of systems theorists only capture the aggregate flow of quantitative parameters, complex systems models consist of unique individual agents capable of learning, planning and symbolic reasoning, capturing the micro-diversity that systems theory did not. Thirdly, Phelan claims that 'complexity' has a different interpretation within the two fields. For systems theorists, "complexity is a function of the number of system components and the amount of interaction between them". In contrast, in complex systems complexity "is something of an illusion—it is something that *emerges* when several agents follow simple rules".

Of these three distinctions, the second is the the most meaningful. The first distinction is a matter of emphasis: both systems theory and complex systems are in the business of both explanation and intervention (consider the interventions recommended by the companies that comprise the Santa Fe Info Mesa). The third distinction may be true for the work of some complex systems theorists – such as Holland [153, 154] and Wolfram [333] – but does not hold for others, such as Bar-Yam [25] and Kauffman [175]. It is more accurate to observe that there are many more ways to think of complexity (algorithmic complexity, behavioural complexity, effective complexity, multiscale complexity, physical complexity and structural complexity, to name a few) in complex systems than the approach initially taken by systems theorists.

The principle difference between complex systems and GST/cybernetics is not the questions that are being asked, but the way in which they are answered. The techniques of complex systems have made substantial ground on the questions that the early systems theorists asked, but did not have the methods to answer. Agent based models, as identified by Phelan, are one component of this expanded toolkit within complex systems. It is only when one delves beyond the level of metaphor that complex systems reveals its contributions to the systems approach.

The analytical techniques within complex systems to understand organisation can be separated by three broad aims: to analyse the patterns, the scales and the dynamics of a system. Patterns that are discovered[12] in measurements or descriptions of a system indicate the formation of structure in either space or time. Kauffman's boolean network models of self-organisation [175], network theory, [35, 322] and computational dynamics [92] are examples of the analysis of pattern formation, where couplings, interactions and correlations respectively between parts of a system give rise to measurable regularities.

The relationship between an observer and a system can be analysed in terms of the scale and scope over which the observer interacts with the system. Multiscale analysis techniques [25, 34, p. 258] provide a method for relating the possible system configurations that are observed at different scales. The development of the renormalisation group[13] [331], and dynamic renormalisation theory to analyse critical phenomena in physics provides an important general method of identifying the

[12]Whether the patterns are a property of the system, the observer, or the relationship between them is a non-trivial question that is intimately related to the theory-ladenness of observation. See Dennett [106] for an insightful discussion on patterns.

[13]Note that there is a connection between the renormalisation group and Simon's nearly decomposable systems [293].

important parameters when the equations describing the system are independent of scale. In such cases, the method of separation of scales does not apply. Renormalisation operates by abstracting interdependencies between fine scale variables into interdependencies between coarse scale variables, and can be applied recursively. Fractals provide another significant set of techniques that enable multiscale analysis in complex systems, which were not available to GST.

Dynamics relates the behaviour of the system to its behaviour in the past. Dynamics can place constraints on future configurations, identify attractors and other patterns that occur in time. The calculus of nonlinear dynamical systems was not invented until after the peak of GST activity, meaning the early systems theorists talked of steady states and equilibria, rather than attractors and bifurcations. The dynamics of evolution and more generally adaptation (which includes learning as well as evolution) are of particular interest, as the only processes that are known to generate multiscale complexity (ie. increase organisation over multiple scales). Advances in biology since GST include a much better understanding of group selection pressures in evolution; swarm intelligence in social insect populations; and the operation of the adaptive immune system; all of which complex systems has drawn on and contributed to.

As well as analytical advances, the impact of the Information Technology revolution on complex systems can hardly be understated. Von Neumann's original automata were computed with pencil and paper. In contrast, Barrett's TRANSIM Agent Based Model (ABM) (see for example [75]), running on Los Alamos National Laboratory supercomputers, has simulated the individual decisions of an entire city's motorists. Currently, there are ABM platforms capable of simulating over a million non-trivial agents, such as MASON [214]. At its best, agent based modelling can combine all three threads of the complex systems approach to understanding organisation, simulating *adaptive pattern formation over multiple scales*. While the computational power and relative simplicity of ABMs can be abused, acting as a substitute for thought[14], it has provided complex systems with a unique approach to model-building that stands in contrast to aggregate models, such as those used in system dynamics.

An example of an aggregate model that has been widely used in defence is the Lanchester differential equation for combat attrition, which relates the mass and effective firing rate of one side to the rate of attrition of the opponent (see Equation (7.1)). Rather than directly guess the functional form of aggregate attrition rates, the agent based approach models the individual agents, and the aggregate behaviour is computed from the 'bottom up'. If the individual behaviour is easier to model than the aggregate behaviour, ABMs can eliminate the need to make simplifying assumptions about the macroscopic dynamics, instead allowing macroscopic patterns to 'emerge' from the microscopic rules and interactions [53]. ABMs are also useful when the spatial configuration is relevant[15] (c.f. Angyal above); when

[14]The complex systems community but also the wider scientific community has been particularly vocal in its criticism of Wolfram's [333] book on this account – see for example [134, 172, 189, 191, 279].

[15]If the spatial dimension is abstracted away but the interactions are retained, an ABM essentially becomes a network model, which has a set of analytical techniques that are increasingly popular within complex systems.

autonomy is distributed across many agents; when thresholds, nonlinearities or logical conditions govern individual decision-making; and when the rules change or the agents learn.

Of course, complex systems is not the only field to simulate the behaviour of individual agents in order to predict aggregate behaviour. In contrast to physics-based constructive simulations, the agents in ABMs are relatively abstract. However, the relationships are typically richer, which gives rise to collective properties in addition to the properties of the simple agents. The limitations of ABMs as predictors of certain aspects of social systems has already been discussed in Section 1.3.

3.4 Defining 'System'

> *A System is a set of variables sufficiently isolated to stay discussable while we discuss it.*
>
> W. Ross Ashby

The historical approach to understanding systems provided a rich context within which individual contributions to systems theory could be assessed. In this section, my aim is to distill this into a concise definition, which captures the essence of the systems approach. Every definition must choose a position along a tradeoff between generality and depth of insight: as Boulding [55] notes, "we always pay for generality by sacrificing content, and all we can say about practically everything is almost nothing." In contrast to Definition 3.1, my definition of 'system' will not be a catch-all container. As a relatively strong definition of system, it will exclude some entities that are legitimately systems in the informal sense – such as closed systems – but which are not the focal subject of the systems approach. Recalling that the focal subject is "organised complexity", this excludes entities with trivial structure or trivial behaviour, such as a contained gas, or well mixed solution of chemicals at equilibrium. The important points to consider from the previous section are:

1. Systems are an idealisation;
2. Systems have multiple components;
3. The components are interdependent;
4. Systems are organised;
5. Systems have emergent properties;
6. Systems have a boundary;
7. Systems are enduring;
8. Systems effect and are affected by their environment;
9. Systems exhibit feedback; and
10. Systems have non-trivial behaviour.

Recall Burke's Definition 2.7 of a system as "an idealisation of an entity as a complex whole." This is an extremely compact definition that incorporates the first seven aspects of systems listed above. Importantly, it explicitly recognises that as an *idealisation*, a system abstracts away many details of an entity (the map is not the territory). A system is not a physical object, it is a representation that stands in for an entity, and it is always constructed and used by an agent (or observer). The use of the word *entity* indicates that the subject of the idealisation may equally be abstract or substantial. The word *complex* implies *multiple components*, *organised* in a non-aggregative arrangement. Also, complexity implies *interdependence* [25, p. 12]. The system *boundary*, its *enduring* nature, and *emergent properties* in addition to the aggregate properties of the parts are all implied by the term *whole* in Burke's definition.

However, the final three aspects of systems are not directly implied by this definition. Eight requires a system to be open. There is a subtle point here that requires explication, due to the fact that the system boundaries are chosen by the observer. Any open system can always be reframed as closed by expanding the system boundaries to include its environment. When applied recursively this method achieves closure with certainty, since the Universe is closed to the exchange of energy, matter and information. However, if for every possible system/environment partitioning the system is closed, then its long term behaviour will be trivial. Consequently, a more accurate way to state this requirement is that there exists at least one system/environment partition such that the system is open.

Together, eight and nine imply ten, since an open system requires more information than the initial conditions of the system to predict its future behaviour, and combinations of positive and negative feedback between the system and its environment are the source of *non-trivial behaviour*. This implication was made precise by von Foerster [313], who distinguished between trivial and non-trivial machines. The former have invariant input-output mappings and determinate behaviour, whereas the latter have input-output relationships that are determined by the machine's previous output. Non-trivial machines may be deterministic, yet are still unpredictable. According to this distinction, *feedback* is necessary and sufficient for non-trivial machines. I now present my definition of system for this thesis, which incorporates all ten aspects of a systems approach:
Definition 3.4 (System (7))**.** *A system is a representation of an entity as a complex whole open to feedback from its environment.*

I define a system as a representation rather than an idealisation, because even though the meaning is similar, as Chapter 2 demonstrated, representation has a well defined theoretical basis. This definition provides the meaning of system for the remainder of this thesis. It is too restrictive as a commonsense conception of system, but it provides insight on both the nature of systemic problems and the value of the systems approach. Because a system is a representation, it can stand in for the entity it represents. Systems are useful idealisations that complement the traditional abstractions within scientific disciplines, such as the point masses, frictionless surfaces, and massless springs of physics. All representations are based on simplifying assumptions, and as a consequence there are limits to their application.

This connection can help to clarify when the systems approach should apply. By rephrasing the list of systems aspects above, one can derive a set of simplifying assumptions of non-systemic approaches which deserve careful examination. This line of reasoning follows closely from Churchman's question, and has been explored elsewhere. For example, Bar-Yam [25, pp. 91-95] takes this approach to contrast complex systems with the assumptions of thermodynamics. Weinberg [324] also provides an extensive discussion of the simplifying assumptions of science, and the importance in systems science of understanding the limits of applicability of simple models. When one or more assumptions are violated, then the systems approach may provide useful insights and more appropriate models (but never complete understanding). The assumptions of non-systemic approaches include:

1. The system is closed;
2. Averaging over time and across individuals is valid;
3. Superpositionality holds;
4. Space can be ignored;
5. Local structure is smooth;
6. Different levels are independent;
7. Control is centralised; and
8. Causation is linear.

The assumption of a closed system enables the design of reproducible experiments with predictable outcomes from the same initial conditions, and the powerful inevitability of increasing entropy obtains, meaning dynamics are irrelevant to long-term behaviour and equilibrium states dominate. However, under this assumption only a limited number of properties of open systems far from equilibrium can be explained. Classical statistics affords great simplification for modelling systems with minimally biased estimators that average over time series and across individual members of an ensemble. However, averaging over time destroys information about temporal structure, while averaging over individuals hides the unique properties of the individual.

Superpositionality means that the whole is the sum of the parts: interactions within the system can be evaluated by calculating the sum of the pairwise interactions between components. This greatly simplifies the modelling process, and is consistent with Occam's Razor, because it assumes the minimal additional information about the system of interest. However, it assumes commutativity and that only first-order dyadic interactions exist. Therefore, superpositionality only strictly applies to aggregates, not systems where the structure of interactions affects the system's properties, and where the forces on a component have indirect effects. Similarly, assuming that spatial arrangement can be ignored simplifies the analysis of well-mixed systems, but if there is significant spatial structure, the assumption of global mixing may give misleading results, as local interactions depart from the average behaviour given by uniformly mixed interactions.

When a system's behaviour is fairly trivial, then one can safely assume that at finer scales, the local structure becomes smooth [25, p. 258], and that different levels can be analysed in isolation. The former assumption allows the application of calculus, while the latter allows the technique of separation of scales, which averages over fast processes and fixes the slow processes to analyse the intermediate dynamics [25, p. 94]. These assumptions do not hold universally. If infinitesimals diverge or correlations occur over all scales – both of which occur at a second order phase transition – new modelling techniques are required.

Assuming control is centralised improves modelling efficiency: most components can be ignored because they are inert in the control process, and control becomes sequential, synchronous and linear. However, if autonomy is distributed across many system components, its behaviour may be qualitatively different to the centralised model. Typically, a system with distributed control may be expected to be less efficient but more robust and adaptive. Finally, the assumption that causation is always best modelled by a linear chain between effects and prior causes, as viewed by Laplace's demon, may not always hold. An example von Foerster [315] gave in an interview with Yveline Rey is how the process of a person switching on a light can be much more efficiently described as how causes in the future, "to have the room lit", determine actions in the present, "turn the switch now". It also explains why if turning the switch on does not illuminate the room because the globe is blown, subsequent actions find an alternative path to the same end, such as "draw the curtains".

This list of assumptions is not complete, but a large proportion of systems research can be motivated in terms of developing alternative techniques for when these assumptions do not hold. Alternatively, systems theory may generalise alternative techniques when they arise within more specialised disciplines, seeking to organise available techniques not by discipline, but by their assumptions about the system's structure and dynamics. Thus, a good deal of attention is paid to surfacing modelling assumptions and discussing limits of applicability within any systems approach. However, where GST found it difficult to say much about systems in general, contemporary systems approaches have had more success in identifying significant features of classes of systems that share certain properties (e.g. chaotic systems), or specific mechanisms for their generation (e.g. genetic algorithms) [293]. The GST ambition to unify science has been supplanted by the narrower but more realistic task of forging interdisciplinary links between specific models within the disciplines.

3.5 Conclusion

The systems movement is an attempt to understand organisation, a concept that is trivialised in mechanism and remains un-analysed in holism. Organisation is a compound concept that incorporates both structure and dynamics. The result of over half a century of systems thinking is not a general theory of organisation, but a loosely connected set of techniques, where each technique contributes some insight on the temporal and spatial structures of organised complexity. Because no single technique provides a complete understanding of organisation, knowledge of the limits of applicability of individual techniques is central to any systems approach.

Complex systems is arguably the most active contemporary field of systems research. It follows in the spirit of general systems theory and cybernetics, although it is less concerned with rhetoric on the unity of science as the development of quantitative models that have interdisciplinary application. Systems analysis and systems engineering have longer and much richer experience with real world systems applications, but constitute a comparatively superficial commitment to systems theory. All of these approaches can be broadly characterised as hard approaches, concerned with the exact scientific application of mathematical models.

The deepest divide between contemporary systems approaches occurs between hard and soft methods. Soft systems methodologies take a pluralist stance, where a systems model is taken to say more about the modeller and their assumptions than the system of interest itself. Advocates of soft systems approaches claim that hard approaches to social systems can be dangerous, because they do not account for the special nature of self-conscious and free-willed humans. Due to the deep philosophical differences between hard and soft systems approaches, there exists little constructive dialog between these communities.

A system is a representation of an entity as a complex whole open to feedback from its environment. The utility of a systems approach derives from the critical examination of simplifying assumptions. This helps to make explicit the associated limits of applicability, such that revision of the appropriate assumptions can extend the application of scientific model-building. The revisions apply in general to reductionist assumptions that wholes do not have properties apart from the properties of their components, and in particular to linear thinking about causation, composition and control.

Chapter 4

Conceptual Analysis

This chapter presents the method of conceptual analysis for complex systems design. Conceptual analysis is a middle path between model-based approaches and metaphorical approaches to complex systems design. It aims to extend the applicability of theorems and insights from complex systems to problems that may be soft, data-poor and social.

4.1 Introduction

Chapter 3 described a number of systems approaches. It was suggested that the most significant difference between the many systems approaches is captured by the hard/soft dichotomy, because it is a metaphysical choice between belief in the one and the many. Consequently, there is a culture gap between hard and soft systems researchers. Hard systems science concentrates on developing mathematical or computer models. Given a problem, a quantitative model is built, the model is analysed and this justifies intervention and design of the system of interest. In contrast, soft systems thinking works within a less structured framework that surfaces assumptions and "root metaphors" of different perspectives. Problems are resolved and decisions are made by finding a strategy that is mutually desirable from multiple perspectives.

Roughly speaking, hard systems approaches dominate every domain except for social problems. Practical applications of hard techniques (especially systems analysis) to social problems have been strongly criticised [81, 157, 208]. The development of a model-based approach to interventions against terrorist attacks was found to be problematic in Chapter 1. Soft systems methodology is a response to perceived failures of hard systems in this domain. Although the results from soft systems have generally been more satisfactory for social problems [225], the current methodology also represents a missed opportunity. By eschewing quantitative methods, soft systems adds very little theoretical insight for the problem owners. Moreover, by emphasising the equality of all perspectives, it dilutes the influence that deep understanding of the system can have. Its strength is bringing structure to the tacit knowledge embedded in organisations, rather than contributing any unique knowledge of systems.

A significant depth of systems knowledge now exists within complex systems. One view of that knowledge from the perspective of information theory will be provided in Chapter 5. The principle research question of this thesis is how it is possible to apply this knowledge without building explicit models of the system of interest. The attempt to answer this question begins in this chapter by advancing a middle path between hard and soft systems research.

I will refer to this approach as conceptual analysis, which is a topic in Yaneer Bar-Yam's course on complex systems. Bar-Yam describes conceptual analysis as the use of complex systems theory at a conceptual level in order to distinguish mountains from molehills. This is precisely the intent behind the applications I will develop in Chapters 7 and 8. The theory is used as a 'model' to represent a system of interest, but it is different to the model-based approaches to complex systems where explicit quantitative models are simulated or solved analytically. More formally, I will work under the following definition:

Definition 4.1 (Multidisciplinary Conceptual Analysis). *An approach to real world problem situations, which draws on theorems, models and exact results from both disciplinary and interdisciplinary science, in order to construct low resolution, conceptual representations that facilitate interventions and evolutionary change within the problem situation.*

Note that multidisciplinary conceptual analysis has a different meaning to the term 'conceptual analysis' in analytic philosophy, which denotes an *a priori* definition of precisely what concepts and words represent. Currently, the most comprehensive collection of social problems that have been addressed using conceptual analysis in the complex systems sense can be found in *Making Things Work: Solving Complex Problems in a Complex World* [34]. Bar-Yam provides a simple yet deep introduction to the theory of complex systems and includes eight real world applications of conceptual analysis. However, the legitimacy of the conceptual analysis approach, its boundaries and its defining characteristics are never explicitly addressed.

This chapter draws on the philosophy of representation and systems developed above to offer broad guidance for conceptual analysis. This guidance is not intended to be prescriptive: conceptual analysis is an approach rather than a methodology. It is also not a framework: the framework of complex systems theory is developed in the second part of my thesis in Chapters 5 and 6. However, it is profitable to articulate the general aims and utility of conceptual analysis in comparison to model-based approaches. Section 4.2 identifies the main characteristics of conceptual analysis. Section 4.3 then raises three issues that must be considered in the application of conceptual analysis: the problems of standards, inconsistency and justification.

4.1.1 Novel contributions

I am the sole author of this chapter, which has not been published elsewhere. My original contributions are:

1. An explicit account of the method of conceptual analysis; and
2. Addressing the issues of standards, inconsistency and justification for multidisciplinary conceptual analysis.

4.2 Multidisciplinary Conceptual Analysis

For systems theory to have an impact, it must ultimately influence real world decisions. If systems theory only interprets the world without changing it, then it is essentially irrelevant. There is a spectrum[1] of actions, from intervention to design, that exert increasing control over a system of interest. Complete control is possible for relatively simple systems, such as the engineering design of a discrete product by a single firm. In more messy situations, such as nation-building to establish viable peace, design is infeasible and even intervention becomes problematic. Conceptual analysis is an approach that links complex systems theory with real world action across the spectrum of intervention to design.

The three distinguishing characteristics of conceptual analysis are:

1. It draws on the exact results of complex systems;

[1]The reader should be aware that this concept of a spectrum of influence assumes the systems theory is developed externally from the system of interest. This is a reasonable assumption in most cases.

2. It does not build explicit quantified models of the system of interest; and
3. It is multidisciplinary.

By "exact results", I mean theorems and symbolic models – the mathematics of complex systems. Mathematics studies the structure of abstractions of quantity, symmetry, change and logic, which roughly correspond to the primary branches of algebra, geometry, analysis and foundations of mathematics. The intention of a systems approach is to explain organisation, which is evident through structure in the configuration and dynamics of systems. Consequently, it is inevitable that mathematics – the systematic study of structure – should play an essential role within complex systems. This role is not to 'mirror' system behaviour, but to provide deep insights on how structure affects function.

The mathematical theorems and agent based models of complex systems can always be interpreted as formal systems. The great strength of formal systems as representations in the sciences is that they generate theorems that are logical consequences of the set of axiomatic assumptions (noting that initial conditions may also be expressed as axiomatic theorems). In other words, mathematics is the only way to ensure that inferences are correct, given a set of explicit assumptions. Ashby felt this gave mathematical explanations a privileged status: "Having experienced the confusion that tends to arise whenever we try to relate cerebral mechanisms to observed behavior, I made it my aim to accept nothing that could not be stated in mathematical form, for only this language can one be sure, during one's progress, that one is not unconsciously changing the meaning of terms, or adding assumptions, or otherwise drifting towards confusion" [290]. This feature of formal systems underwrites the confidence of hard scientists in the language of mathematics. Most importantly, it focuses debate on the axiomatic assumptions, since any disagreement between a theorem and empirical experiment is traceable to the initial assumptions.

In Section 3.4, it was shown how the systems approach can be characterised as the critical appraisal and revision of scientific assumptions. In this view, it is strongly complementary to the use of mathematics, which derives the necessary consequences of a set of assumptions. It is perfectly reasonable to assume that the behaviour of any system can be profitably analysed by the critical use of formal systems representations, since as Section 3.4 showed, systems are enduring, and consequently have regularities that can be represented. The main problem arises when social factors are forcibly quantified, and rules governing the dynamics are explicitly specified. There are many cases when this is either not possible or not appropriate. What insights can exact techniques offer in these situations?

When explicit quantitative models are produced in systems design, it is easy to understand how they function as representations that shape behaviour (ie. shape the final design). An illustrative description of the process is as follows. The modeller – who in this case is the agent in Definition 2.2 – firstly quantifies their purpose using an objective function $f(\cdot)$. Secondly, the modeller identifies the variables in the system of interest that affect the value of $f(\cdot)$, which can be partitioned by variables $x \in X$ the modeller can influence, and variables $y \in Y$ where significant influence is not possible. The modeller sets the value of y based on measurements or expectations of the system of interest, then explores the space

X for good values of $f(x, y)$ until a termination criteria is reached. The model provides a useful substitute for physical designs, since in general moving in model design space is much cheaper than moving in physical design space. Finally, the behaviour of the modeller is influenced by the model, because the best solution x^* at the time of termination of movement in model space guides the physical design that is implemented. An equally clear-cut recommendation for intervention can be made on the basis of quantitative models.

The case for conceptual models differs significantly. The modeller does not need to reduce their understanding of objectives $f(\cdot)$ or environmental factors y to quantitative measurements. This freedom comes at the cost of much lower resolution. The relative fitness of two similar configurations may be unclear. Further, a conceptual model is not intended to identify a point solution x^*. Instead, the model distinguishes between volumes of design space, one or more of which may be recommended for certain contexts (ie. for a set of environmental conditions $y \in \mathcal{S}_Y$). A conceptual model is used to understand the coarse features of good design, in order to identify the most significant parameters. This helps to avoid "subsystematisation" [158, p. 160], when more detailed models of easily quantified subsystems are optimised at the expense of total system performance. Conceptual models direct the modeller towards broad classes of designs or intervention strategies, rather than specifying a point design.

Most disciplines have had the debate over quantitative versus qualitative models. The trend towards increasing empirical grounding in almost all disciplines [185, p. 198] has forced the question with respect to almost every subject domain. At the same time, this trend has tended to add currency to quantitative approaches and marginalised more qualitative approaches.

The debate within economics is especially interesting, because economics has historically placed a greater emphasis on mathematical models of social phenomena than the other social sciences. As a result, economists have a depth of knowledge in both mathematics and social problems – the domain where hard systems methods have had the least success.

A particularly eloquent discussion on this topic took place over several dozen letters between Keynes and Harrod in 1938 [46]. The discussion concerned the applicability of the methods of the natural sciences within economics, and in particular the value of populating economic models with quantitative statistical data. Keynes and Harrod agreed that the work of Shultz, attempting to turn economics into a "pseudo-natural-science", was more or less a waste of time. However, they disagreed on the merits of quantitative research as exemplified by Tinbergen's research on business cycles. Harrod's opinion was that

> Tinbergen may be doing very valuable work, in trying to reduce this part of theory to quantitative terms ... I begin to feel that the time has come when I ought to soil my fingers by doing some of this sort of statistical work myself or supervise others in the doing of it.

Keynes' reply sheds light on why quantifying parameters in social models may not always be beneficial.

> In chemistry and physics and other natural sciences the object of exper-

> iment is to fill in the actual values of the various quantities and factors appearing in an equation or a formula; and the work when done is once and for all. In economics that is not the case, and to convert a model into a quantitative formula is to destroy its usefulness as an instrument of thought...
>
> One has to be constantly on guard against treating the material as constant and homogeneous in the same way that the material of the other sciences, in spite of its complexity, is constant and homogeneous. It is as though the fall of the apple to the ground depended on the apple's motives, on whether it is worth while falling to the ground, and whether the ground wanted the apple to fall, and on mistaken calculations on the part of the apple as to how far it was from the centre of the earth.
>
> But do not be reluctant to soil your hands, as you call it. I think it is most important. The specialist in the manufacture of models will not be successful unless he is constantly correcting his judgment by intimate and messy acquaintance with the facts to which his model has to be applied.

This correspondence ran in parallel to a conversation between Tinbergen and Harrod, which resulted in Tinbergen rewriting the introduction to his book to accommodate Keynes' critique. At the heart of the matter is Keynes concern that *to convert a model into a quantitative formula is to destroy its usefulness as an instrument of thought*. Keynes also touches on the limitation of modelling social phenomena that can be found as a disclaimer on any financial product disclosure statement: past performance is no guarantee of future results. This distinction between models in economics and the natural sciences changes the way that models are used. In economics, according to Keynes, the "object of statistical study is not so much to fill in missing variables with a view to prediction, as to test the relevance and validity of the model." Empirical test is still important, but it Keynes wanted to emphasise that it did not fulfil the same function for social phenomena. Soiling one's hands could aid with choosing between competing models and improving understanding of the relevance of a model, but it should not be expected to converge towards a predictive tool where the empirical data directly supports intervention.

Just as the exact theories in economics can provide a basis for decision-making from market design to government regulatory interventions – even when they are not furnished with statistical data – so too can the exact theories of complex systems inform design and intervention. In the spirit of Keynes' philosophy, conceptual analysis uses models to change the way the modeller thinks of the problem situation, rather than as containers for statistically valid measurements of quantities that operate under assumptions of stationarity and context independence.

4.3 Problems for Multidisciplinary Conceptual Analysis

The main difference between Keynes' approach and conceptual analysis is that conceptual analysis is multidisciplinary. Multidisciplinarity is important, because it is not appropriate to demarcate between simple and complex systems: systems are a way of representing the world, they are not the furniture of the world. Consequently, the theories of complex systems are entangled with disciplinary knowledge, and applications of complex systems must be able to also draw on disciplinary discourse as required. However, this has three implications that need to be addressed: the problems of standards, inconsistency and justification.

The twentieth century phenomena of increasingly differentiated disciplines is a more or less inevitable consequence of the growth of science, and the changes to how science is funded. For problems that admit to partitioning by subject, the disciplinary framework allows standards, conventions and aspirations for researchers to be shared, which become embodied in communities of practice and professional societies. According to Klein [184] in her book on interdisciplinarity, a discipline signifies "a stable epistemic community and agreement upon what constitutes excellence in a field". Chapter 3 introduced the systems movement, which was an experiment in establishing an interdisciplinary professional society. As a central figure of general systems theory, Boulding ([56] as cited in [184]) noted that interdisciplinarity "easily becomes the undisciplined if there is no organized payoff for the constant critical selection of its ideas, theories and data". The observation is even more pertinent to multidisciplinarity. It begs the question: how can the standards of disciplinary research be maintained outside of disciplines?

One answer is that they cannot. However, because the purpose of multidisciplinarity is fundamentally different, it is not appropriate to measure value of multidisciplinary research directly against disciplinary standards. Nevertheless, some structure must be imposed on activity outside of disciplinary frameworks. This was the motivation behind Boulding's elaboration of the skeleton of science discussed in Section 3.3.1. Determining the appropriate structure for non-disciplinary discourse is a substantial issue that has not been satisfactorily addressed to date. Klein [184, pp. 12-13] cites three reasons why interdisciplinary discourse is problematic. They are confusion over definitions; widespread unfamiliarity with interdisciplinary scholarship; and the dispersion of discourse across general, professional, academic, government, and industrial literatures. The paradox that non-disciplinary discourse faces is that should it become structured, professionalised and specialised, it will be yet another discipline. Only by adding structure can non-disciplinary research consistently raise its standards, but it will have lost its defining character in the process, and more importantly will have walled itself off from the disciplines it seeks to integrate. Perhaps the only way out of this paradox is to specialise, but to do so in a way that runs orthogonal to the disciplinary decomposition. In this way, multidisciplinary discourse can be bounded and structured, but it can also confront problems that no other individual discipline can address.

The second problem for multidisciplinary research is the inconsistency of assumptions between disciplines. Because of this, the uncritical concatenation of theories from different disciplines produces *non sequiturs*. Literary theorist Fish [124] is critical toward the very idea of interdisciplinarity because he believes it invariably mixes incompatible sources. Discussing examples of this practice, Fish concludes that meaningful research can only occur within narrow disciplinary confines.

> The unconcern these [interdisciplinary] essays display is with the claims made by the theorists cited, claims that arise from the particular problems they set out to solve, problems that are urgent and perspicuous in the context of some specific project... But if you have your eye on a larger horizon—a horizon so large that it barely knows boundaries, never mind laws of entailment—almost anything you come across will seem relevant and capable of being plugged in unproblematically.

However, it is inevitable that there will exist messy real world problems that are broader than any disciplinary framework. By denying the possibility of interdisciplinary (or multidisciplinary) research, Fish is denying that theory can be systematically applied to such problems. This view seems too extreme, although the logical problems associated with integrating often eclectic collections of disciplinary theory (also raised by Hoos [157, 158] and Lilienfeld [208]) need to be addressed.

To avoid drawing invalid conclusions from integrating over incompatible sources, there are two requirements for a multidisciplinary approach. Firstly, the assumptions of any model need to be considered, including implicit assumptions that derive from its disciplinary origin. This aspect of systems theory has already been elaborated in the previous chapter. Secondly, components of a multidisciplinary approach need to be examined together, and any inconsistencies must be explicitly addressed. Identifying the reasons for inconsistencies between disciplinary theories and providing potential resolutions can be a major contribution of a multidisciplinary approach.

The third problem for a multidisciplinary approach is the problem of justifying the use of theories that cannot be directly empirically tested. Because multidisciplinary theories are necessarily broader than a single discipline, they are less readily tested. They must firstly be supplemented with the context provided by a discipline before an experiment can be designed to collect empirical evidence. Consider Bak's [23] theory of self-organised criticality. The idea that many natural systems self-organise towards a critical point is an exceptionally difficult theory to test. As many examples can be found that refute the theory as support it. Consequently, self-organised criticality has itself remained critically poised on the edge between acceptance as a useful contribution to complex systems and rejection as an unsubstantiated and overly broad claim. Ambiguity about its domain of applicability, and the need for significant furnishing with disciplinary context have both contributed to uncertainty of the value of self-organised criticality.

Computer modelling can be useful here, since artificial worlds can be arbitrarily abstract, unlike physical experiments. However, while simulations can provide a stringent test the on the logical validity of a theory, they tend to be a more elastic test of empirical validity. The flexibility of arbitrary abstractness also means that important physical details may be abstracted away, giving misleading support to general theories. Fortunately for conceptual analysis, this issue is less problematic, because the approach operates with a coarse enough resolution that the need for empirical testing is largely circumvented. Another way to reduce the risk of utilising theories that are not empirically testable is to introduce them at a sufficiently fine scale, which then reduces the scale of both negative and positive consequences. Feedback can then inhibit counterproductive changes and reinforce useful changes. When empirical evidence is unavailable, small fractions of the real world system can be used as a laboratory. This is the role of test pilots, and more generally of pilot programs to introduce cultural change, education initiatives and even new television shows into existing systems.

A final aspect of multidisciplinary research is relevant to all three issues raised above. The value of a multidisciplinary approach should ultimately be judged not whether it is correct, but whether it is useful, either directly or indirectly. The problems of standards, inconsistency and justification are all analytical concerns. However, there will always be very real limits to the omniscience of *a priori* analysis. These limits are implied by the distinction between representations and representation bearers (Section 2.3), are emphasised by Churchman's question about securing improvement for the whole system (Section 3.3), and will be demonstrated for the problem of predicting emergence (Section 6.5). Together, these concepts point to the conclusion that investing too heavily in analytic justification for action can be counterproductive. An alternative perspective recasts the role of modelling as generating novel variation in potentially fertile regions of design space. The merits of novel designs can then be measured in operational environments. Low resolution conceptual models allow new designs to be produced with a relatively small investment of time and resources, which need to be tested in context. Selection pressure can then operate directly on physical designs to reinforce useful variation. The use of real evolutionary pressure in context to improve engineering designs is described in detail in [27]. Rather than seeking to establish an analytical guarantee for successful design, this use of multidiscplinary conceptual analysis focusses on improving the design over time through situated learning and evolution.

4.4 Conclusion

This chapter presented the case for multidisciplinary conceptual analysis. It was argued that even in problems where the environment is non-stationary, fitness is not easily quantifiable and social factors are significant, the exact theories and theorems of complex systems are relevant. In conceptual analysis, low resolution models are constructed that draw on both complex systems and disciplinary theory. The purpose of conceptual analysis is to identify promising classes of significantly novel interventions or system designs to seed evolutionary change in real world systems. As a multidisciplinary approach, conceptual analysis needs to pay special attention to the assumptions underlying the theory it draws upon, address inconsistencies in cross-disciplinary models, and make small scale changes to systems in context that utilise evolutionary mechanisms to improve system performance over time. Conceptual analysis can apply in contexts where traditional quantitative model-based applications of complex systems are inappropriate, and provide insights that soft systems methodologies do not reveal.

Part II

Theory

Chapter 5

The Information Theory of Complex Systems

Complex systems as a discipline currently works within a loosely defined framework of concepts such as complexity, self-organisation, emergence and adaptation. The fuzziness of complex systems definitions is complicated by the unclear relation among these central processes: does self-organisation emerge or does it set the preconditions for emergence? Does complexity arise by adaptation or is complexity necessary for adaptation to arise? The inevitable consequence of this lack of clarity is miscommunication. We propose a set of concepts, together with their information-theoretic interpretations, which can be used as a dictionary of complex systems discourse. Our hope is that this information-theoretic framework may clarify communication between practitioners, and provide new insights into the field.

5.1 Introduction

Chapter 3 introduced complex systems, which studies general phenomena of systems comprised of many elements interacting in a non-trivial fashion. Fuzzy quantifiers like ‘many’ and ‘non-trivial’ are inevitable. ‘Many’ implies a number large enough so that no individual component or feature predominates the dynamics of the system, but not so large that features are completely irrelevant. Interactions need to be non-trivial so that the degrees of freedom are suitably reduced, but not constraining to the point that the arising structure possesses no further degree of freedom. Crudely put, systems with a huge number of components interacting trivially are explained by statistical mechanics, and systems with precisely defined and constrained interactions are the concern of fields like chemistry and engineering. In so far as the domain of complex systems overlaps these fields, it contributes insights when the classical assumptions are violated. In Figure 3.1, Weinberg’s depiction of types of systems shows this region of “organised complexity”. However, such simple depictions can give the misleading impression that there is a hard discontinuity between organised complexity and the contrasting regions of organised simplicity and disorganised complexity. In reality, the boundaries between these regions contains systems where two alternative idealisations both apply to some degree.

Due to the difficulty of precisely defining the domain of complex systems, it is unsurprising this vagueness extends to other aspects of the discipline, which notably lacks a standard formal framework for analysis. There are a number of reasons for this. Because complex systems is broader than physics, biology, sociology, ecology, or economics, its foundations cannot be reduced to a single discipline. Furthermore, systems which lie in the gap between the ‘very large’ and the ‘fairly small’ do not submit meekly to traditional mathematical modelling techniques or frameworks.

Initially setting aside the requirement for formal definitions, we can summarise our general understating of complex systems dynamics as follows:

1. Complex systems are <u>open systems</u>, and receive a regular supply of energy, information, and/or matter from the environment;
2. A medium sized ensemble of individual components interact, which gives rise to non-trivial <u>behaviour</u>;
3. The non-trivial interactions result in internal constraints, leading to <u>symmetry breaking</u> in the behaviour of the individual components, from which coordinated global behaviour arises;
4. The system is now more <u>organised</u> than it was before; since no central director nor any explicit instruction template was followed, we say that the system has <u>self-organised</u>;
5. This coordination can express itself as <u>patterns</u> detectable by an external observer or as structures that convey new properties to the system itself. New behaviours <u>emerge</u> from the system;
6. Coordination and emergent properties may arise in response to environ-

mental pressure, in which case we can say the system and its parts display adaptation;

7. When adaptation occurs across generations at a population level we say that the system evolved;
8. Coordinated emergent properties give rise to larger scale effects. These interdependent sets of components with emergent properties can be observed as coherent entities at lower resolution than is needed to observe the components. The system can be identified as a novel unit of its own and can interact with other systems/processes expressing themselves at the same scale. This becomes a building block for new iterations and the cycle can repeat from 1. above, now at a larger scale.

The process outlined above is not too contentious, but does not yet address how and why each step occurs. Consequently, even when we can recognise the process in a phenomenological way, we can not understand it, intervene with confidence, or engineer for it. A scientific explanation must address the deeper issue of the mechanisms that give rise to the phenomenology.

Even worse than the fuzziness and absence of deep understanding already described, is when the above terms are used interchangeably in the literature. The danger of not making clear distinctions in complex systems is incoherence. To have any hope of coherent communication, it is necessary to unravel the knot of assumptions and circular definitions that are often left unexamined.

Here we suggest a set of working definitions for the above concepts, essentially a dictionary for complex systems discourse. Our purpose is not to be prescriptive, but to propose a baseline for shared agreement, to facilitate communication between scientists and practitioners in the field. We would like to prevent the situation in which a scientist talks of emergence and this is conflated with self-organisation.

Part I provided a high level overview of complex systems, and established why and in what circumstances it applies to real world problems. This chapter begins the task of providing the detailed framework for complex systems theory. For this purpose we chose an information-theoretic framework. There are a number of reasons for this choice:

- Shannon's [287] theory of communication had an enormous influence on the early systems work in cybernetics and general systems theory;
- A considerable body of work in complex systems has been cast into information theory, as pioneered by the Santa Fe Institute, and we borrow heavily from this tradition;
- It provides a well developed theoretical basis for our discussion;
- It provides definitions which can be formulated mathematically;
- It provides a range of computational tools, so for simple cases quantitative measures can be computed.

Nevertheless, we believe that the concepts should also be accessible to disciplines which often operate beyond the application of such a strong mathematical and computational framework, like biology, sociology and ecology. Consequently, for each concept we provide a 'plain English' interpretation, which is intended to enable communication across fields. In this spirit, we have also attempted to depict all of the concepts in a single graphic to provide an intuitive understanding of the significant interrelationships.

5.1.1 Novel contributions

This chapter is based on:

[252] M. Prokopenko, F. Boschetti, and A. J. Ryan. An Information-theoretic Primer on Complexity, Self-organisation and Emergence. *Submitted to Complexity*, 2007.

Our original contributions are:

1. A comprehensive survey of information theoretic interpretations of complexity, the edge of chaos, self-organisation, emergence, adaptation, evolution and self-referentiality;
2. A definition for each concept, both in the mathematics of information theory and in plain English; and
3. Insights into the relationships between complex systems concepts.

My role in this work was:

1. Leading the section on adaptation and evolution;
2. Leading the discussion and conclusions;
3. Helping with the writing of each section; and
4. Developing the graphical interpretation of the relationship between the concepts.

There are some differences between [252] and the version I present here. I have rewritten the section on emergence to be more consistent with Chapter 6 and expanded the discussion in many sections.

5.2 An information-theoretical approach

Information theory was originally developed by Shannon [287] for the reliable transmission of information from a source X to a receiver Y over noisy communication channels, and includes the fundamental concepts of source coding and channel coding.

Source coding quantifies the average number of bits needed to represent the result of an uncertain event. Intuitively, it measures one's freedom of choice in selecting a message which captures information in the source X faithfully i.e. without information loss. Another interpretation is the number of possible distinctions that are required to distinguish between every potentially different configuration of X. This quantity is known as *(information) entropy.* In the simplest cases, when all configurations are equiprobable, the amount of information can be measured by the logarithm (base 2) of the number of configurations. The entropy is a precise measure of the amount of freedom of choice (or the degree of randomness) contained in the process – a process with many possible states has high entropy. Formally, given a probability distribution P over a (discrete) random variable X, the entropy is defined by

$$H(X) = -\sum_{X} p(x) \log p(x) \ , \tag{5.1}$$

and is the only measure satisfying three required properties:

- Changing the value of one of the probabilities by a small amount changes the entropy by a small amount;
- If all the choices are equally likely, then entropy is maximal; and
- Entropy is independent of how the process is divided into parts.

The joint entropy of two (discrete) random variables X and Y is defined as the entropy of the joint distribution of X and Y:

$$H(X,Y) = -\sum_{X,Y} p(x,y) \log p(x,y) \ , \tag{5.2}$$

Mutual information $I(X;Y)$ is defined as the amount of information (in bits) that the received signal Y contains about the transmitted signal X, on average:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = H(Y) - H(Y|X) \ , \tag{5.3}$$

where $H(Y|X)$ is the conditional entropy of Y given X, also called the equivocation of Y about X. If X and Y are independent, the joint entropy is simply the sum of their individual entropies, and the mutual information is zero. Informally, we can state that

$$\begin{aligned} \text{mutual information} &= \text{receiver's diversity} \\ &- \text{equivocation of receiver about source} \end{aligned} \tag{5.4}$$

Equivocation of Y about X may also be interpreted as non-assortativeness between Y and X: the degree of having no correlation in either a positive or negative way. This interpretation is particularly useful in dealing with assortative, disassortative, and non-assortative networks. While this 'plain English' definition may still seem obscure, mutual information is simply the average reduction in uncertainty that knowledge of X provides about the outcome of Y. For example, if $H(Y) = H(Y|X)$, then $I(X;Y) = 0$. Because the uncertainty about Y does not decrease when X is known, X provides no information about the value of Y, and so their mutual information is 0.

Channel coding establishes that reliable communication is possible over noisy channels if the rate of communication is below a certain threshold called the channel capacity. Channel capacity is defined as the maximum mutual information for the channel over all possible distributions of the transmitted signal X.

In the remainder of this chapter, we intend to point out how different concepts in complex systems can be interpreted via simple information-theoretic relationships. In particular, when suitable information channels are identified, the rest is often a matter of computation – the computation of "diversity" and "equivocation". The choice of channels is typically a task for modellers, which we will demonstrate with numerous examples.

There are other mathematical approaches, such as non-linear time series analysis, chaos theory, etc., that also provide insights into the concepts of complex systems. However, these approaches are outside the scope of this chapter. It is possible that information theory has not been widely used in applied studies of complex systems because the literature can appear impenetrable to the uninitiated: there is a high mathematical barrier to entry. We would like to clarify its applicability and illustrate how different information channels can be identified and used.

5.3 Complexity

It is an intuitive notion that certain processes and systems are harder to describe than others. Complexity tries to capture this difficulty in terms of the amount of information needed for the description, the time it takes to carry out the description, the size of the system, the number of components in the system, the number of conflicting constraints, the number of dimensions needed to embed the system dynamics, and related ideas. For a much broader survey of measures of complexity than can be supplied here, see `http://bruce.edmonds.name/combib`, which contains 386 references with distinct definitions of complexity.

Here we propose to adopt a definition of complexity as the amount of information needed to describe a process, a system or an object. This definition is computable (at least in one of its forms), is observer-independent (once resolution is defined), it applies to both data and models [54] and provides a framework within which self-organisation and emergence can also be consistently defined.

5.3.1 Concept

Algorithmic Complexity The original formulation can be traced back to Solomonoff, Kolmogorov and Chaitin, who developed independently what is today known as Kolmogorov-Chaitin or algorithmic complexity [79]. Given an entity (this could be a data set or an image, but the idea can be extended to material objects and also to life forms) the algorithmic complexity is defined as the length (in bits of information) of the shortest program (computer model) which can describe the entity. According to this definition a simple periodic object (a sine function for example) is not complex, since we can store a sample of the period and write a program which repeatedly outputs it, thereby reconstructing the original data set with a very small program. At the opposite end of the spectrum, an object with no internal structure cannot be described in any meaningful way but by storing every feature, since we cannot rely on any shared structure for a shorter description. It follows that a random object has maximum complexity, since the shortest program able to reconstruct it needs to store the object itself. Note that this provides a definition of randomness for individual objects, as having a structure which can not be compressed in any meaningful sense. In this way, algorithmic complexity provides deeper insight than probability theory, where randomness is defined with reference to an ensemble. A nice property of this definition is that it does not depend on what language we use to write the program since it can be shown that descriptions using different languages differ only by constants.

Statistical Complexity A clear disadvantage of the algorithmic complexity is that it can not be computed exactly but only approximated from above[1]. A common criticism of this definition is that there are problems for which associating randomness to maximum complexity seems counter-intuitive. Imagine you throw a cup of rice to the floor and you want to describe the spatial distribution of the grains. In most cases you do not need to be concerned with storing the position of each individual grain; the realisation that the distribution is structure-less and that predicting the exact position of a specific grain (given the location of other grains) is impossible is probably all you need to know. And this piece of information is very simple (and short) to store. There are applications for which our intuition suggests that both strictly periodic and totally random sequences should share low complexity.

One definition which addresses this concern is the statistical complexity [97], which attempts to measure the size of the minimum program able to reproduce the statistically significant features of the entity under analysis. In the rice pattern mentioned above, there is no statistical difference in the probability of finding a grain at different positions and the resulting statistical complexity is zero. Along with our previous non-mathematical analogy we could say that for many purposes the actions of a mentally disabled patient, taking apparently random actions at any time can be described very briefly by just noticing the lack of apparent intentionality or purpose in the patient's behaviour.

Apart from implementation details, the conceptual difference between algorithmic and statistical complexity lies in how randomness is treated. Essentially, the algorithmic complexity implies a deterministic description of an object (it defines the information content of an individual sequence), while the statistical complexity implies a statistical description (it refers to an ensemble of sequences generated by a certain source) [52, 137]. As suggested by Boffetta et al. [52], which of these approaches is more suitable is problem-specific. In the previous analogy, the details of the behaviour of a mentally-disabled patient maybe of crucial importance to a psychiatrist, who may thus prefer an algorithmic approach, but less relevant to a more superficial observer, for whom a statistical description would suffice.

Excess entropy and predictive information As pointed out by Bialek et al. [47], our intuitive notion of complexity corresponds to statements about the underlying process, and not directly to Kolmogorov complexity. A dynamic process with an unpredictable and random output (large algorithmic complexity) may be as trivial as the dynamics producing predictable constant outputs (small algorithmic complexity) – while "really complex processes lie somewhere in between". Interestingly, however, the two extreme cases share one feature: the entropy of the

[1]See the Chaitin theorem [78].

output strings "either is a fixed constant or grows exactly linearly with the length of the strings", and corrections to the asymptotic behaviour do not grow with the size of the data set. Grassberger [137] identified the slow approach of the entropy to its extensive limit as a sign of complexity. Thus, subextensive components – which grow with time less rapidly than a linear function – are of special interest. Bialek et al. [47] observe that the subextensive components of entropy identified by Grassberger determine precisely the information available for making predictions – e.g. the complexity in a time series can be related to the components which are "useful" or "meaningful" for prediction. We shall refer to this as *predictive information*. Revisiting the two extreme cases, they note that "it only takes a fixed number of bits to code either a call to a random number generator or to a constant function" – in other words, a model description *relevant to prediction* is compact in both cases.

The predictive information is also referred to as excess entropy [94, 96], stored information [289], effective measure complexity [119, 137, 210], complexity [16, 207], and has a number of interpretations.

5.3.2 Information-theoretic interpretation

Predictive information

In order to estimate the relevance to prediction, two distributions over a stream of data x are considered: a prior probability distribution for the futures, $\mathrm{P}(x_{future})$, and a more tightly concentrated distribution of futures conditional on the past data, $\mathrm{P}(x_{future}|x_{past})$, and define their average ratio

$$I_{pred}(T, T') = \left\langle log_2 \frac{\mathrm{P}(x_{future}|x_{past})}{\mathrm{P}(x_{future})} \right\rangle, \tag{5.5}$$

where $\langle \cdots \rangle$ denotes an average over the joint distribution of the past and the future, $\mathrm{P}(x_{future}|x_{past})$, T is the length of the observed data stream in the past, and T' is the length of the data stream that will be observed in the future. This average predictive information captures the reduction of entropy (in Shannon's sense) by quantifying the information that the past provides about the future:

$I_{pred}(T, T') = H(T') - H(T'|T)$ or informally,

$$\begin{aligned} \text{predictive information} = {} & \text{total uncertainty about the future} - \\ & \text{uncertainty about the future, given the past} \end{aligned} \tag{5.6}$$

The predictive information is always positive and grows with time less rapidly than a linear function, being subextensive. It provides a universal answer to the question of how much is there to learn about the underlying pattern in a data stream: $I_{pred}(T, T')$ may either stay finite, or grow infinitely with time. If it stays finite, this means that no matter how long we observe we gain only a finite amount of information about the future: e.g. it is possible to completely predict dynamics of periodic regular processes after their period is identified. For some irregular processes the best predictions may depend only on the immediate past (e.g. a Markov process, or in general, a system far away from phase transitions and/or symmetry breaking) – and in these cases $I_{pred}(T, T')$ is also small and is bound by the logarithm of the number of accessible states: the systems with more states and longer memories have larger values of predictive information [47]. If $I_{pred}(T, T')$ diverges and optimal predictions are influenced by events in the arbitrarily distant past, then the rate of growth may be slow (logarithmic) or fast (sublinear power). If the data allows us to learn a model with a finite number of parameters or a set of underlying rules describable by a finite number of parameters, then $I_{pred}(T, T')$ grows logarithmically with a coefficient that counts the dimensionality of the model space (i.e. the number of parameters). Sublinear power-law growth may be associated with infinite parameter (or nonparametric) models such as continuous functions with some regularisation (e.g. smoothness constraints) [48].

Statistical complexity

The statistical complexity is calculated by reconstructing a minimal model, which contains the collection of all situations which share a similar specific probabilistic future – the causal states – and measuring the entropy of the probability distribution of the states. The description of an algorithm which achieves such reconstruction and calculates the statistical complexity for 1D time series can be found in [284] and for 2D time series in [283].

In general, the predictive information is related to the statistical complexity such that
$I_{pred}(T, T') \leq C_\mu$, which is the minimum average amount of memory needed to statistically reproduce the configuration ensemble to which the sequence belongs [306] – both predictive information and statistical complexity are measured in bits.

Excess entropy

Excess entropy is defined as a measure of the total apparent memory or structure in a source [95]:

$$E = \sum_{L=1}^{\infty} (h_\mu(L) - h_\mu), \tag{5.7}$$

where $h_\mu(L) = H(L) - H(L-1), L \geq 1$, for the entropy $H(L)$ of length-L sequences or blocks, and $h_\mu = \lim_{h\to\infty} \frac{H(L)}{L}$ is the source entropy rate – also known as per-symbol entropy, the thermodynamic entropy density, Kolmogorov-Sinai entropy, metric entropy, etc. The length-L entropy rate $h_\mu(L)$ is the average uncertainty about the L^{th} symbol, provided the $(L-1)$ previous ones are given [52]. As noted by Crutchfield and Feldman [95], the length-L approximation $h_\mu(L)$ typically overestimates the entropy rate h_μ at finite L, and each difference $\lfloor h_\mu(L) - h_\mu \rfloor$ is the difference between the entropy rate conditioned on L measurements and the entropy rate conditioned on an infinite number of measurements – it estimates the information-carrying capacity in the L-blocks that is not actually random, but is due instead to correlations, and can be interpreted as the local (i.e. L-dependent) predictability [115]. The total sum of these local over-estimates is the excess entropy or intrinsic redundancy in the source. Thus, the excess entropy measures the amount of apparent randomness at small L values that is "explained away" by considering correlations over larger and larger blocks. Importantly, Crutchfield and Feldman [95] demonstrated that the excess entropy E can also be seen as either:

1. the mutual information between the source's past and the future – exactly the predictive information $I_{pred}(T, T')$, if T and T' are semi-infinite, or

2. the subextensive part of entropy $H(L) = E + h_\mu L$, as $L \to \infty$.

It was also shown that only the first interpretation holds in 2-dimensional systems [123].

Revisiting Equation (5.6), we may point out that the total uncertainty $H(T')$ can be thought of as structural diversity of the underlying process. Similarly the conditional uncertainty $H(T'|T)$ can be related to structural non-conformity or equivocation within the process. That is, a degree of non-assortativeness between the past and the future, or between components of the process in general. This analogy creates an alternative intuitive representation, which will be analysed in section 5.4:

$$\text{predictive information} = \text{excess entropy} = \text{diversity} - \text{non-assortativeness} \quad (5.8)$$

5.3.3 Example – Thue-Morse process

Predictive information grows logarithmically for infinite-memory Thue-Morse sequences $\sigma^k(s)$ which contain two units 0 and 1, and can be obtained by the substitution rules $\sigma^k(0) = 01$ and $\sigma^k(1) = 10$ (e.g. $\sigma^2(1) = 1001$, etc.). Such a process needs an infinite amount of memory to maintain its aperiodicity [95], and hence, its past provides an ever-increasing predictive information about its future.

An even faster rate of growth is also possible – typically, this happens in problems where predictability over long scales is "governed by a progressively more detailed description" as more data are observed [47]. This essentially produces an increase in the number of causal states, used by Crutchfield in defining statistical complexity.

5.3.4 Example – periodic vs random processes

The source entropy rate h_μ captures the irreducible randomness produced by a source after all correlations are taken into account [95]:

- $h_\mu = 0$ for periodic processes and even for deterministic processes with infinite-memory (e.g. Thue-Morse process) which do not have an internal source of randomness, and

- $h_\mu > 0$ for irreducibly unpredictable processes, e.g. independent identically distributed (IID) processes which have no temporal memory and no complexity, as well as Markov processes (both deterministic and nondeterministic), and infinitary processes (e.g. positive-entropy-rate variations on the Thue-Morse process).

The excess entropy, or predictive information, is a better measure of complexity as it increases with the amount of structure or memory within a process:

- E is finite for both periodic and random processes (e.g. it is zero for an IID process) – its value can be used as a relative measure: a larger period results in higher E, as a longer past needs to be observed before we can estimate the finite predictive information;
- E diverges logarithmically for complex processes due to an infinite memory (e.g. Thue-Morse process) – again, its value can be used as a relative measure estimating a number of parameters or rules in the underlying model;
- E exhibits a sublinear power law divergence for complex processes due to a nonparametric model of the underlying process (e.g. a continuous function with smoothness constraints) – here, the relative measure is the number of different parameter-estimation scales growing in proportion to the number of taken samples (e.g. the number of bins used in a histogram approximating the distribution of a random variable), i.e. a progressively more detailed description is required.

5.3.5 Summary

While entropy rate is a good identifier of intrinsic randomness, it suffers from the same drawbacks as the Kolmogorov-Chaitin (KC) complexity, to which it is strongly related. To reiterate, the KC complexity of an object is the length of the minimal Universal Turing Machine (UTM) program needed to reproduce it. The entropy rate h_μ is equal to the average length (per variable) of the minimal program that, when run, will cause an UTM to produce a typical configuration and then halt [91, 122, 206].

The relationships $I_{pred}(T, T') = E$ and $E \leq C_\mu$, suggest a very intuitive interpretation:

$$\begin{aligned} \text{predictive information} &= \\ \text{richness of structure} &\leq \text{statistical complexity} \\ &= \text{memory for optimal predictions} \end{aligned} \tag{5.9}$$

where the latter is defined as the entropy of causal states – all histories that have the same conditional distribution of futures [97, 282]. The causal states provide an optimal description of a system's dynamics in the sense that these states make as good a prediction as the histories themselves. The inequality in Equation (5.9) means that the memory needed to perform an optimal prediction of the future configurations cannot be lower than the mutual information between the past and future themselves [122]: this relationship reflects the fact that the causal states are a reconstruction of the hidden, effective states of the process.

Specifying how the memory within a process is organised cannot be done within the framework of information theory [95], and a more structural approach based on the theory of computation must be used – this leads (via causal states) to ε-machines and statistical complexity C_μ.

5.4 Edge of Chaos

According to the analysis presented in the preceding section, statistical complexity is small at both extremes (complete order and complete randomness), and is maximal in the region somewhere between the extremes. Moreover, in some "intermediate" cases, the complexity is infinite, and may be divergent at different rates. A few natural questions then arise:

- Where does complexity become maximal;
- How can we precisely identify this region;
- What happens to information dynamics within or near this region; and
- How are such dynamics are related to self-organisation, adaptation, evolution, and emergence in general.

5.4.1 Concept

Cellular automata (CA) are discrete spatially-extended dynamical systems that have been used as models of many computational, physical and biological processes [226], and we shall use this abstraction to illustrate the edge of chaos phenomenon[2]. Langton [196] examined behaviour of CA in terms of the parameter λ – the fraction of rules with a given property in the rule table (e.g. the fraction of nonzero transitions). Varying the parameter within its range between ordered and chaotic extremes, he identified, for intermediate values of λ, an increase in the mutual information $I(A;B)$ between a cell and itself at the next time-step. An intriguing feature is that the average mutual information has a distinct peak at the transition point – an indication of a phase transition from "order" to "chaos" in CA. This study introduced the edge of chaos – the region where the CA behaviour shifts from the ordered regimes towards chaotic regimes, effectively approaching random dynamics.

5.4.2 Information-theoretic interpretation

A rule-space of 1-dimensional CA can be characterised with the Shannon entropy of the rules' frequency distribution [335]. More precisely, given a rule-table (the rules that define a CA), the input-entropy at time step t is defined as

$$S^t = -\sum_{i=1}^{m} \frac{Q_i^t}{n} \log \frac{Q_i^t}{n} \tag{5.10}$$

where m is the number of rules, n is the number of cells (system size), and Q_i^t is the number of times the rule i was used at time t across the CA. The input-entropy settles to fairly low values for ordered dynamics, but fluctuates irregularly

[2]We would like to point out that the edge of chaos hypothesis is far from being accepted unanimously. The edge of chaos concept is also distinct from low dimensional chaos in dynamical systems, characterised by non-periodicity with few degrees of freedom. Nevertheless some connections have been made – see subsection 5.4.3 for a brief description.

within a narrow high band for chaotic dynamics. For the complex CA, order and chaos may predominate at different times causing the entropy to vary. A measure of the variability of the input-entropy curve is its variance or standard deviation, calculated over time. Wuensche [335] has convincingly demonstrated that only complex dynamics exhibits high variance of input-entropy, leading to automatic classification of the rule-space. Importantly, the peak of input-entropy variance points to a phase transition as well, indicating the edge of chaos.

Similarly, Wolfram [332] classified cellular automata rules qualitatively – according to their asymptotic behaviour: class I (homogeneity); class II (periodicity); class III (chaos); class IV (complexity). The first class consists of CA that, after a finite number of time steps, produce a unique, homogeneous state (analogous to "fixed points" in phase space). From almost all initial states, such behaviour completely destroys any information on the initial state, i.e. complete prediction is trivial and complexity is low. The second class contains automata which generate a set of either stable or periodic structures (typically having small periods – analogous to "limit cycles" in phase space), where each region of the final state depends only on a finite region of the initial state, i.e. information contained within a small region in the initial state thus suffices to predict the form of a region in the final state. The third class includes infinite CA producing aperiodic ("chaotic") spatiotemporal patterns from almost all possible initial states – the effects of changes in the initial state almost always propagate forever at a finite speed, and a particular region depends on a region of the initial state of an ever-increasing size (analogous to "strange attractors" in phase space). While any prediction of the "final" state requires complete knowledge of the initial state, the regions are indistinguishable statistically as they possess no structure, and therefore the statistical complexity is low. The fourth class deals with automata that generate patterns continuously changing over an unbounded transient.

5.4.3 Example – universal computation

The fourth class of CA contains members at the edge of chaos (in the critical region) that were also shown to be capable of universal computation [332]. They support three basic operations (information storage, transmission, and modification) through static, propagating and interacting structures (blinkers, gliders, collisions). Importantly, the patterns produced along the transient are different in terms of generated structure, and in fact, their structural *variability* is highest among all four classes – i.e. the predictive information and the complexity of the class IV automata are highest. Casti [75] made an analogy between the complex (class IV) automata and quasi-periodic orbits in phase space, while pursuing deeper interconnections between CA, dynamical systems, Turing Machines, and formal logic systems – in particular, the complex automata producing edge of chaos dynamics were related to formal systems with undecidable statements (Gödel's Theorem).

5.4.4 Example – graph connectivity

Graph connectivity can be analysed in terms of the size of the largest connected subgraph (LCS) and its standard deviation obtained across an ensemble of graphs, as suggested by Random Graph Theory [118]. In particular, critical changes occur in connectivity of a directed graph as the number of edges increases: the size of the LCS rapidly increases as well and fills most of the graph, while the variance in the size of the LCS reaches a maximum at some critical point before decreasing. In other words, variability within the ensemble of graphs grows as graphs become increasingly different in terms of their structure.

An information-theoretic representation can subsume this graph-theoretic model. For example, a feasible average measure of a complex network's heterogeneity is given by the entropy of a network defined through the link distribution. The latter can be defined via the simple degree distribution, the probability P_k of having a node with k links, or via the remaining degree distribution q_k: "the number of edges leaving the vertex other than the one we arrived along" [296]. The remaining degree distribution is useful in analysing how assortative, disassortative or non-assortative is the network. Assortative mixing (AM) in Graph Theory is the extent to which high-degree nodes connect to other high degree nodes [233]. In disassortative mixing (DM), high-degree nodes are connected to low-degree ones. Both AM and DM networks are contrasted with non-assortative mixing (NM), where one cannot establish any preferential connection between nodes on the basis of degree.

Importantly, the conditional entropy $H(q|q')$ may estimate spurious correlations in the network created by connecting the vertices with dissimilar degrees – this noise affects the overall diversity or the average uncertainty of the network, but does not contribute to the amount of information (correlation) within it. Using the joint probability of connected pairs $q_{k,k'}$, one may calculate the amount of correlation between vertices in the graph via the mutual information measure, the information transfer, as

$$I(q) = H(q) - H(q|q') = \sum_{k=1}^{m} \sum_{k'=1}^{m} q_{k,k'} \log \frac{q_{k,k'}}{q_k q_{k'}}. \tag{5.11}$$

Informally,

$$\begin{aligned} \text{transfer within the network} &= \text{diversity in the network} \\ &- \text{assortative noise in the network structure} \end{aligned} \tag{5.12}$$

This interpretation is analogous to the one suggested by the Equations (5.4) and (5.8), assuming that assortative noise is the non-assortative extent to which the preferential (either AM or DM) connections are obscured. In general, the mutual information $I(q)$ is a better, more generic measure of dependence: correlation functions, like the variance in the size of the LCS, "measure linear relations whereas mutual information measures the general dependence and is thus a less biased statistic" [296]. At the edge of chaos, the information transfer within the network attains its maximum. This case, however, is not easily achievable as observed by Solé and Valverde. In fact, the cases where entropy $H(q)$ and noise $H(q|q')$ are approximately equal, are most typical, suggesting that some intrinsic constraints dominate the search-space of possible network configurations.

5.4.5 Summary

The main edge of chaos hypothesis asserts that when biological systems must perform complex computation for survival, the process of evolution under natural selection tends to select systems near a phase transition between ordered and chaotic behaviour [226] – i.e. the complex dynamic regime acts as an evolutionary attractor in the state space. The analysis summarised in this section emphasises the role of structural variability (equivalent to predictive information) in complex systems – at the edge of chaos the behaviour is potentially structurally richer and therefore, has a higher predictive capacity.

5.5 Self-Organisation

Three ideas are implicit in the word self-organisation: a) the organisation in terms of global implicit coordination; b) the dynamics implicit in progressing in time from a not (or less) organised to an organised state; and c) the spontaneous arising of such dynamics. To avoid semantic traps, it is important to notice that the word 'spontaneous' should not be taken literally; we deal with open systems, exchanging energy, matter and/or information with the environment and made up of components whose properties and behaviours are defined prior to the organisation itself. The 'self' prefix merely indicates that no centralised ordering or external agent/template explicitly guides the dynamics. It is thus necessary to define what is meant by 'organisation' and how its arising or increase can be detected.

5.5.1 Concept

A commonly held view is that organisation entails an increase in complexity, since both are related to interdependence, and both are minimal for the case of independence. Unfortunately the lack of agreement on complexity means we are no closer to understanding organisation. De Wolf and Holvoet [103] define complexity as a measure of redundancy or structure in the system. The concept can be made more formal by adopting the statistical complexity described above as a measure of complexity, as demonstrated in Shalizi [280] and Shalizi et al. [285]. This definition offers several of the advantages of the Computational Mechanics approach: it is computable and observer independent. Also, it captures the intuitive notion that the more a system self-organises, the more behaviours it can display, the more effort is needed to describe its dynamics. Importantly, this needs to be seen in a statistical perspective; while a disorganised system may potentially display a larger number of actual configurations, the distinction among several of them may not be statistically significant. Adopting the statistical complexity allows us to focus on the system configurations which are statistically different (causal states) for the purpose at hand. We thus have a measure which is based only on the internal dynamics of the system (and consequently is observer-independent) but which can be tuned according to the purpose of the analysis.

For an alternative definition of self-organisation based on thermodynamics as studied by Prigogine, and the distinction between self-organisation and the related concept of self-assembly, we refer the reader to Halley and Winkler [144].

5.5.2 Information-theoretic interpretation

In the scientific literature the concept of self-organisation refers to both living and non-living systems, ranging from physics and chemistry to biology and sociology. Kauffman [176] suggests that the underlying principle of self-organisation is the generation of constraints in the release of energy. According to this view, the constrained release allows for such energy to be controlled and channelled to perform some useful work. This work in turn can be used to build better and more efficient constraints for the release of further energy and so on. This principle is closely related to Kauffman's own definition of life [176]. It helps us to understand why an organised system with effectively less available configurations may behave and look more complex than a disorganised one to which, in principle, more configurations are available. The ability to constrain and control the release of energy may allow a system to display behaviours (reach configurations) which, although possible, would be extremely unlikely in its non-organised state. It is surely possible that 100 parrots move independently to the same location at the same time, but this is far more likely if they fly in a flock. A limited number of coordinated behaviours become implementable because of self-organisation, which would be extremely unlikely to arise in the midst of a nearly infinite number of disorganised configurations. The ability to constrain the release of energy thus provides the self-organised system with behaviours that can be selectively chosen for successful adaptation.

However, Halley and Winkler [144] correctly point out that attention should paid to how self-organisation is treated if we want the concept to apply equally to both living and non-living systems. For example, while it is tempting to consider adaptation as a guiding process for self-organisation, it then makes it hard to use the same definition of self-organisation for non-living systems.

Recently, Correia [89] analysed self-organisation motivated by embodied systems, i.e. physical systems situated in the real world, and established four fundamental properties of self-organisation: no external control, an increase in order, *robustness*[3], and interaction. All of these properties are easily interpretable in terms of information transfer. Firstly, the absence of external control may correspond to 'spontaneous' arising of information dynamics without any flow of information into the self-organising system. Secondly, an increase in order or complexity reflects simply that the predictive information is increased within the system or its specific part:

$$I_{pred}([t_1 - T, t_1], [t_1, t_1 + T']) < I_{pred}([t_2 - T, t_2], [t_2, t_2 + T']), \text{ and} \tag{5.13}$$

$$C_\mu^{System}(t_2) > C_\mu^{System}(t_1), \text{ for } t_2 > t_1 \text{ and positive } T \text{ and } T', \tag{5.14}$$

[3] Although Correia refers to this as adaptability, according to the concepts in this chapter he in fact defines robustness. This is an example of exactly the kind of issue we hope to avoid by developing this dictionary.

where $C_\mu^{System}(t)$ is the statistical complexity at time t. In general, however, we believe that one may relax the distinction between these two requirements and demand only that in a self-organising system, the amount of information flowing from the outside $I_{Outside}$ is strictly less than the change in the predictive information's gain:

$$I_{Outside} < I_{pred}([t_2 - T, t_2], [t_2, t_2 + T']) - I_{pred}([t_1 - T, t_1], [t_1, t_1 + T']), \text{ or } \quad (5.15)$$

$$C_\mu^{Outside} < C_\mu^{System}(t_2) - C_\mu^{System}(t_1), \quad (5.16)$$

where $C_\mu^{Outside}$ is the complexity of the contribution from the outside.

In general, a system is robust if it continues to function in the face of perturbations [317]. Information-theoretically, robustness of a self-organising system to perturbations means that it may interleave stages of an increased information transfer within some channels (dominant patterns are being exploited) with periods of decreased information transfer (alternative patterns are being explored).

The interaction property is described by Correia [89] as follows: "minimisation of local conflicts produces global optimal self-organisation, which is evolutionarily stable." Minimisation of local conflicts, however, is only one aspect, captured in Equations (5.4), (5.8), and (5.12) as equivocation or non-assortativeness, and should be generally complemented by maximising diversity within the system.

5.5.3 Example – self-organising traffic

In the context of pedestrian traffic, Correia [89] argues that it can be shown that the "global efficiency of opposite pedestrian traffic is maximised when interaction rate is locally minimised for each component. When this happens two separate lanes form, one in each direction. The minimisation of interactions follows directly from maximising the average velocity in the desired direction." In other words, the division into lanes results from maximising velocity (an overall objective or fitness), which in turn supports minimisation of conflicts.

Another example is provided by ants: "Food transport is done via a trail, which is an organised behaviour with a certain complexity. Nevertheless, a small percentage of ants keeps exploring the surroundings and if a new food source is discovered a new trail is established, thereby dividing the workers by the trails [161] and increasing complexity" [89]. Here, the division into trails is again related to an increase in fitness and complexity.

These two examples demonstrate that when local conflicts are minimised, the degree of coupling among the components (i.e. interaction) increases and the information flows easier, thus increasing the predictive information. This means that not only the overall diversity of a system is important (more lanes or trails), but the interplay among different channels (the assortative noise within the system, the conflicts) is crucial as well.

5.5.4 Example – self-regulating morphogenetic processes

Modelling information processing in bio-systems is concerned with the nature of biological information and the ways in which it is processed in biological and artificial cells and tissues. There are a number of models involving self-organising or self-regulating processes within complex biological systems. For example, some morphogenetic models explore the requirements for spontaneous formation of regular and stable structures out of more or less homogeneous cell aggregations or cell sheets – i.e. the creation of a new form apparently without a detailed external blueprint. In particular, the biomechanical model of Beloussov and Grabovsky [41] reproduces two main categories of patterns, stationary cell domains and running waves. It interplays short- and long-range factors in a "causal chain" moving from the state of an initial homogeneous cell layer towards the complicated shapes of embryonic rudiments: "the local perturbations act in ensemble, rather than in a mosaic, one-to-one manner"[4]. The model involves a measure of the cells' mechanosensitivity (i.e. their ability to transform external mechanical stresses into active mechanochemical reactions), and changes in the tangential pressure due to cell extensions and contractions. An increase in mechanosensitivity is related to the formation of regular stationary structures and propagating waves, travelling across embryonic tissues, while an increase of the limits of tangential extension is essentially destructive, deteriorating regular structures and generating irregular waves. This interplay between two forces may be interpreted in information-theoretic terms: the information transfer is optimal when the diversity in mechanosensitivity is large, and the assortative noise (conflicts) inducing tensions is minimal. Spontaneous formation occurs, thus, within an optimal range of the information transfer and not at either of these extremes.

5.5.5 Example – self-controlling neural automata

One possible pathway towards an optimal solution in the information-theoretic search-space is explored in the context of neural networks with dynamic synapses. Cortes et al. [90] studied neural automata (neurobiologically inspired cellular automata) which exhibit "chaotic itinerancy among the different stored patterns or memories". In other words, activity-dependent synaptic fluctuations ("noise") explore the search-space by continuously destabilising a current attractor and inducing random hopping to other possible attractors. The complexity and chaoticity of the resulting dynamics (hopping) depends on the intensity of the synaptic "noise" and on the number of the network nodes that are synchronously updated (we note again the two forces involved in the information transfer – diversity and non-assortativeness). Cortes et al. [90] utilised a quantitative measure – the entropy of neural activity over time (computed in frequency-domain), and related varying values of synaptic noise parameter F to different regimes of chaoticity. Decreasing entropy for mid-range values of F indicates a tendency towards regularisation or smaller chaoticity. Importantly, hopping controlled by synaptic noise

[4]We would like to point out that in this example, as well as in many others, there are external initial conditions, and this requires a precise estimation of the external influence $C_\mu^{Outside}$, contrasted with the increase in complexity $C_\mu^{System}(t_2) - C_\mu^{System}(t_1)$ during morphogenesis.

may occur autonomously, without the need for external stimuli.

5.5.6 Example – self-organising locomotion

The internal channels through which information flows within the system are observer-independent, but different observers may select different channels for a specific analysis. For example, let us consider a modular robotic system modelling a multi-segment snake-like (salamander) organism, with actuators ("muscles") attached to individual segments ("vertebrae"). A particular side-winding locomotion emerges as a result of individual control actions when the actuators are coupled within the system and follow specific evolved rules [253, 254]. There is no global coordinator component in the evolved system, and it can be shown that the amount of predictive information between groups of actuators grows as the modular robot starts to move across the terrain – that is, the distributed actuators become more coupled when a coordinated side-winding locomotion is dominant. Faced with obstacles, the robot temporarily loses the side-winding pattern: the modules become less organised, the strength of their coupling is decreased, and rather than exploiting the dominant pattern, the robot explores various alternatives. Such exploration temporarily decreases self-organisation, i.e. the predictive information within the system. When the obstacles are avoided, the modules "rediscover" the dominant side-winding pattern by themselves, recovering the previous level of predictive information and manifesting again the ability to self-organise without any global controller. Of course, the "magic" of this self-organisation is explained by properties defined *a priori*: the rules employed by the biologically-inspired actuators have been obtained by a genetic programming algorithm, while the biological counterpart (the rattlesnake *Crotalus cerastes*) naturally evolved over a long time. Our point is simply that we can measure the dynamics of predictive information and statistical complexity as it presents itself within the channels of interest.

In summary, the fundamental properties of self-organisation are immediately related to information dynamics, and can be studied in precise information-theoretic terms when the appropriate channels are identified.

5.6 Emergence

Traditionally, the idea of emergence has been intimately associated with a layered view of nature. Emergence explains the existence of irreducible wholes at levels above fundamental physics. Examples that have been considered include such grand events as the emergence of classical reality from quantum mechanics [180], of life from non-living chemistry [100], and of mind from non-conscious neurons [305]. As was discussed in Chapter 3, emergence was initially proposed as a middle ground between the mechanists and the vitalists, by claiming that emergent properties of systems were consistent with, but not reducible to the properties of their physical components. The emergent properties require the use of new concepts at a 'higher level', which are meaningless at the level below [81]. For example, quarks, molecules, cells, humans is a potential sequence of wholes, each of which (above quarks) have emergent properties not applicable to the level below. It can also be noticed that scientific models of these wholes become progressively less detailed as the level increases. The resolution decreases, by decreasing the granularity of the distinctions that the observer is capable of discerning. This is partly because it is impractical to model every subatomic particle in a human, but it is also partly because it is unnecessary: emergence of a whole implies coherence and therefore redundancy.

Attempts to formally address the study of emergence have sprung up at regular intervals in the last century or so (they are reviewed in Chapter 6), under different names, with varied approaches and motivations. Here we will examine Crutchfield's [93], distinction between two phenomena which have been both been discussed as candidates for emergence: pattern formation and 'intrinsic' emergence. Both concepts involve agents forming representations to shape their own behaviour (see Chapter 2).

5.6.1 Concept

Pattern Formation. In pattern formation we imagine an observer trying to 'understand' a process. If the observer detects some patterns (structures) in the system, they can then employ such patterns as tools to simplify their understanding of the system. As an example, a gazelle which learns to correlate hearing a roaring to the presence of a lion, will be able to use it as warning and flee danger. Not being able to detect the pattern 'roaring = lion close by' would require the gazelle to detect more subtle signs, possibly needing to employ more attention and thus more effort. In this setting the observer (gazelle) is 'external' to the system (lion) it needs to analyse and represent in a way that modifies its own behaviour.

Intrinsic emergence. In intrinsic emergence, the observer is 'internal' to the system. Imagine a set of traders in an economy. The traders are locally connected via their trades, but no global information exchange exists. Once the traders identify an 'emergent' feature, say in the stock market, they can employ it to understand and *affect* the functioning of the system itself. The stock market becomes a mean for global information processing, which is performed by the agents (that is, the system itself) to affect their *own* functioning.

5.6.2 Information-theoretic interpretation

Given that a system can be described and studied at different levels, a natural question is "what level should we choose for our analysis"? A reasonable answer could be "the level at which it is easier or more efficient to construct a workable model". This idea has been captured formally by Shalizi [280] in the definition of Efficiency of Prediction. Within a Computational Mechanics [282] framework, Shalizi suggests:

$$e = \frac{E}{C_\mu} \tag{5.17}$$

where e is the Efficiency of Prediction, E is the excess entropy and C_μ the statistical complexity discussed above. The excess entropy can be seen as the mutual information between the past and future of a process, that is, the amount of information observed in the past which can be used to predict the future (i.e. which can be usefully coded in the agent instructions on how to behave in the future). Recalling that the statistical complexity is defined as the amount of information needed to reconstruct a process (that is equivalent to performing an optimal prediction), we can write informally:

$$e = \frac{\text{how much can be predicted}}{\text{how difficult it is to predict}} \tag{5.18}$$

Given two levels of description of the same process, the approach Shalizi suggests is to choose for analysis the level which has larger efficiency of prediction e. At this level, either:

- we can obtain better predictability (understanding) of the system (E is larger), or
- it is much easier to predict because the system is simpler (C_μ is smaller), or
- we may lose a bit of predictability (E is smaller) but at the benefit of much larger gain in simplicity (C_μ is much smaller).

We can notice that this definition applies equally to pattern formation as well as to intrinsic emergence. In the case of pattern formation, we can envisage the scientist trying to determine what level of enquiry will provide a better model. At the level of intrinsic emergence, developing an efficient representation of the environment and of *its own functioning within the environment* gives a selective advantage to the agent, either because it provides for a better model, or because it provides for a similar model at a lower cost, enabling the agent to direct resources towards other activities. This definition means it is possible to make precise and measurable comparison of the efficiency of different models as discussed in subsection 2.4.1.

5.6.3 Example – the emergence of thermodynamics

A canonical example of emergence without self-organisation is described by Shalizi [280]: thermodynamics can emerge from statistical mechanics. The example considers a cubic centimetre of argon, which is conveniently spinless and monoatomic, at standard temperature and pressure, where the gas is sampled every nanosecond. At the micro-mechanical level, and at time intervals of 10^{-9} seconds, the dynamics of the gas are first-order Markovian, so each microstate is a causal state. The thermodynamic entropy (calculated as $6.6 \cdot 10^{20}$ bits) gives the statistical complexity C_μ. The entropy rate h_μ of one cubic centimetre of argon at standard temperature and pressure is quoted to be around $3.3 \cdot 10^{29}$ bits per second, or $3.3 \cdot 10^{20}$ bits per nanosecond. Given the range of interactions $R = 1$ for a first-order Markov process, and the relationship $E = C_\mu - Rh_\mu$ [122], it follows that the efficiency of prediction $e = E/C_\mu$ is about 0.5 at this level. Looking at the macroscopic variables uncovers a dramatically different situation. The statistical complexity C_μ is given by the entropy of the macro-variable energy which is approximately 33.28 bits, while the entropy rate per millisecond is 4.4 bits (i.e. $h_\mu = 4.4 \cdot 10^3$ bits/second). Again, the assumption that the dynamics of the macro-variables are Markovian, and the relationship $E = C_\mu - Rh_\mu$ yield $e = E/C_\mu = 1 - Rh_\mu/C_\mu = 0.87$. If the time-step is a nanosecond, like at the micro–mechanical level, then $e \approx 1$, i.e. the efficiency of prediction approaches maximum. This allows Shalizi to conclude that "almost all of the information needed at the statistical-mechanical level is simply irrelevant thermodynamically", and given the apparent differences in the efficiencies of prediction at two levels, "thermodynamic regularities are emergent phenomena, emerging out of microscopic statistical mechanics" [280].

5.7 Adaptation and Evolution

Adaptation is a process where the behaviour of the system changes such that there is an increase in the fit, and therefore the mutual information, between the system and a potentially complex and non-stationary environment. The environment is treated as a black box, meaning an adaptive system does not need to understand the underlying system dynamics to adapt. Stimulus response interactions provide feedback that modifies an internal model or representation of the environment, which affects the probability of the system taking future actions.

5.7.1 Concept

The three essential functions for an adaptive mechanism are generating variety, observing feedback from interactions with the environment, and selection to reinforce some interactions and inhibit others. Without variation, the system cannot change its behaviour, and therefore it cannot adapt. Without feedback, there is no way for changes in the system to be coupled to the structure of the environment. Without preferential selection for some interactions, changes in behaviour will not be statistically different to a random walk. First order adaptation keeps sense and response options constant and adapts by changing only the probability of future actions. However, adaptation can also be applied to the adaptive mechanism itself [138]. Second order adaptation introduces three new adaptive cycles: one to improve the way variety is generated, another to adapt the way feedback is observed and thirdly an adaptive cycle for the way selection is executed. If an adaptive system contains multiple autonomous agents using second order adaptation, a third order adaptive process can use variation, feedback and selection to change the structure of interactions between agents.

From an information-theoretic perspective, variation decreases the amount of information encoded in the system, while selection acts to increase information. Since adaptation is defined to increase mutual information between a system and its environment, the information loss from variation must be less than the increase in mutual information from selection.

For the case that the system is a single agent with a fixed set of available actions, the environmental feedback is a single real valued reward plus the observed change in state at each time step, and the internal model is an estimate of the future value of each state, this model of first order adaptation reduces to reinforcement learning (see for example [298]).

For the case that the system contains a population where the fitness of successive generations is correlated due to inheritance with variation under selective pressure, the adaptive process reduces to evolution. Evolution is not limited to DNA/RNA based terrestrial organisms, since other entities, including populations of individuals, prions, artificial life programs and culture also meet the criteria for evolution given by Lewontin [205]:

1. Different individuals in a population have different morphologies, physiologies, and behaviors (phenotype variation).
2. Different phenotypes have different rates of survival and reproduction in different environments (differential fitness).
3. There is a correlation between parents and offspring in the contribution of each to future generations (fitness is heritable).

One well-known generalisation of evolution is Dawkins' [102] concept of Universal Dawinism, which extends the Darwinian principles of variation, inheritance and selection to apply to evolutionary change in all open, complex systems.

5.7.2 Information-theoretic interpretation

Adami [3] advocated the view that "evolution increases the amount of information a population harbours about its niche". In particular, he proposed physical complexity – a measure of the amount of information that an organism stores in its genome about the environment in which it evolves. Importantly, physical complexity for a population X (an ensemble of sequences) is defined in relation to a specific environment Z, as mutual information:

$$I(X, Z) = H_{max} - H(X|Z) \tag{5.19}$$

where H_{max} is the entropy in the absence of selection, i.e. the unconditional entropy of a population of sequences, and $H(X|Z)$ is the conditional entropy of X given Z, i.e. the diversity tolerated by selection in the given environment. When selection does not act, no sequence has an advantage over any other, and all sequences are equally probable in ensemble X. Hence, H_{max} is equal to the sequence length. In the presence of selection, the probabilities of finding particular genotypes in the population are highly non-uniform, because most sequences do not fit the particular environment. The difference between the two terms in Equation (5.19) reflects the observation that "If you do not know which system your sequence refers to, then whatever is on it cannot be considered information. Instead, it is potential information (a.k.a. entropy)". In other words, this measure captures the difference between potential and selected (filtered) information:

$$\begin{aligned}\text{physical complexity} = {} & \text{how much data can be stored} - \\ & \text{how much data irrelevant to environment is stored}\end{aligned} \tag{5.20}$$

Comparing this with the information transfer through networks, Equation (5.12), as well as analogous information dynamics Equations (5.4) and (5.8), we can observe a strong similarity: "how much data can be stored" is related to diversity of the network, while “how much data irrelevant to environment is stored” (or “how much conflicting data”) corresponds to assortative noise in the network. In short, natural selection increases physical complexity by the amount of information a population contains about its environment. Adami argued that physical complexity must increase in molecular evolution of asexual organisms in a single niche if the environment does not change, due to natural selection, and that “natural selection can be viewed as a filter, a kind of semipermeable membrane that lets information flow into the genome, but prevents it from flowing out”. In general, information may flow out and it is precisely this dynamic that creates larger feedback loops between the system and its environment.

Adami’s definition of physical complexity can be related back to our earlier definition of predictive information. Adami states that “physical complexity *is* information about the environment that can be used to make predictions about it.” There is an important difference between physical complexity and predictive information, excess entropy and statistical complexity. Whereas the latter three measure correlations within a sequence, physical complexity measures correlation between two sequences representing the system and its environment. However, it is always possible to represent the system and its environment as a single sequence by redefining the system boundary to include the environment, which means that correlations between the system and its environment occur within a sequence. This can then be measured in principle by statistical complexity, provided that correlations within the system, and correlations within the environment are accounted for.

5.7.3 Example – perception-action loops

The information transfer can also be interpreted as the acquisition of information from the environment by a single adapting individual: there is evidence that pushing the information flow to the information-theoretic limit (i.e. maximisation of information transfer) can give rise to intricate behaviour, induce a necessary structure in the system, and ultimately adaptively reshape the system [186, 187]. The central hypothesis of Klyubin et al. is that there exists “a local and universal utility function which may help individuals survive and hence speed up evolution by making the fitness landscape smoother”, while adapting to morphology and ecological niche. The proposed general utility function, *empowerment*, couples the agent’s sensors and actuators via the environment. Empowerment is the perceived amount of influence or control the agent has over the world, and can be seen as the agent’s potential to change the world. It can be measured via the amount of Shannon information that the agent can “inject into” its sensor through the environment, affecting future actions and future perceptions. Such a perception-action loop defines the agent’s actuation channel, and technically empowerment is defined as the capacity of this actuation channel: the maximum mutual information for the channel over all possible distributions of the transmitted signal. “The more of the information can be made to appear in the sensor, the more control or influence the agent has over its sensor” – this is the main motivation for this local and universal utility function [187]. Other examples highlighting the role of information transfer in guiding selection of spatiotemporally stable multi-cellular patterns, well-connected network topologies, and coordinated actuators in a modular robotic system are discussed in [253, 254, 256, 257].

5.8 Self Referentiality

Patterns of intrinsic emergence may form as a result of observer-independent processes – nevertheless, these are patterns only with respect to a certain observation structure, imposed on or selected in the environment *a priori*, and not necessarily by the "observer". An interesting aspect, however, is that the choice of such a structure is not entirely decoupled from the environment. In fact, this choice often depends on the emergent phenomena. So the process can be best characterised in terms of tangled hierarchies exhibiting Strange Loops: "an interaction between levels in which the top level reaches back down towards the bottom level and influences it, while at the same time being itself determined by the bottom level" [152].

5.8.1 Concept

This view involves the concept of *downward causation* [49, 136, 150]: a feature is emergent if it has some sort of causal power on lower level entities. While common views of emergence assume that lower level entities must have an "upward" causation on the emergent features, this approach requires a 2-way causal relation. As an example, we can imagine individuals organising into a community (e.g. the Bar problem [17]). Their actions affect how the community develops (upward causality) and the development of the community itself affects the behaviour and interaction of the individuals (downward causality). An important feature of such systems is that, in the absence of explicit information sharing, each individual may need to "best-guess" what the other individuals may do over time. Systems that fall into this category are sometimes called self-referential problems – situations where the forecasts made by the individuals serve to create the world they are trying to forecast [39, 40]. In these systems, the "best" thing to do depends not on a rational decision model (which does not exist), but on what everyone else is doing, creating a heterogeneous mix of decision strategies that continually co-evolve over time [17, 38, 77]. Even if the individuals are allowed to communicate and share some information, the heterogeneity prevails [116].

5.8.2 Information-theoretic interpretation

A closely related problem is the minority game [80]: a repeated game where N (odd) agents have to choose one out of two alternatives at each time step, and those who happen to be in the minority win. As noted by Batten [39], "seemingly simple at first glance, the game is subtle in the sense that if all players analyse the situation identically, all will choose the same alternative and therefore all will lose". The agents may share information on the outcomes of the game in the past M rounds (memory). In the minority game, there is a second order phase transition in the space of the parameter $\alpha = 2^M/N$, between a symmetric phase (where no past information is available to agents) and an asymmetric phase (where past information is available). An important feature of minority games is the possibility for some agents (risk-averters) to imitate others (risk-takers) in selecting their actions. As observed by Slanina [294], there is an optimal level of imitation, beyond which the whole collective starts to perform worse in terms of the average gain of the agents: "moderate imitation can be beneficiary, while exaggerated one can be harmful". In addition, this optimal level increases with decreasing memory length M: the less information is shared, the more imitation is allowed.

5.8.3 Example – shortest path formation by ants

Another well-known example of downward causation is shortest path formation with ants: ants indirectly interact through changes in their environment by depositing pheromones and forming a pheromone trail. This mechanism for increasing shared information is known as *stigmergy*. The probability that an ant chooses a trail increases with the number of ants that chose the same path in the past – the process is thus characterised by a positive feedback loop (a form of autocatalytic behaviour called *allelomimesis*) [112]. Since the pheromone evaporates (loss of information), the shortest path lasts longer than the alternatives, and the ants have a higher probability of using and reinforcing it. Thus, not only the shortest path emerges as a result of allelomimesis and stigmergy (upward causation), it also influences the movement of the ants because they follow the pheromone trail (downward causation) [103]. Importantly, we again observe diversity in ant behaviours (path explorers and path exploiters).

5.8.4 Summary

The observed diversity in agent behaviours (risk-takers and risk-averters, or explorers and exploiters) creates information dynamics: loss of information, or "forgetting" effect [193], and gain of information, or "information selectivity" [194]. Our conjecture is that self-referentiality eventually forces a split in information dynamics, breaking the cycle between upward and downward causation. As a result of such a break, some form of memory (implicit or explicit) emerges in the system in order to support both diversity and imitation.

In summary, in self-referential systems there is also a relationship between heterogeneity, the level of non-assortative mixing, and the information shared within the system – similar to the information transfer relationships in Equations (5.4), (5.8), (5.12), and (5.20). A precise information-theoretic analysis of this relationship for self-referential systems is a subject of future research (see Parunak et al. [244] and Prokopenko et al. [255] for preliminary analysis).

5.9 Discussion and Conclusions

By studying the processes which result from the local interaction of relatively simple components irrespective of composition, complex systems has accepted the audacious aim of addressing problems which range from physics to biology, sociology and ecology. It is not surprising that a common framework and language which enable practitioners of different field to communicate effectively is still lacking. As a possible contribution to this goal we have proposed a framework within which concepts like complexity, emergence and self-organisation can be described, and most importantly, distinguished.

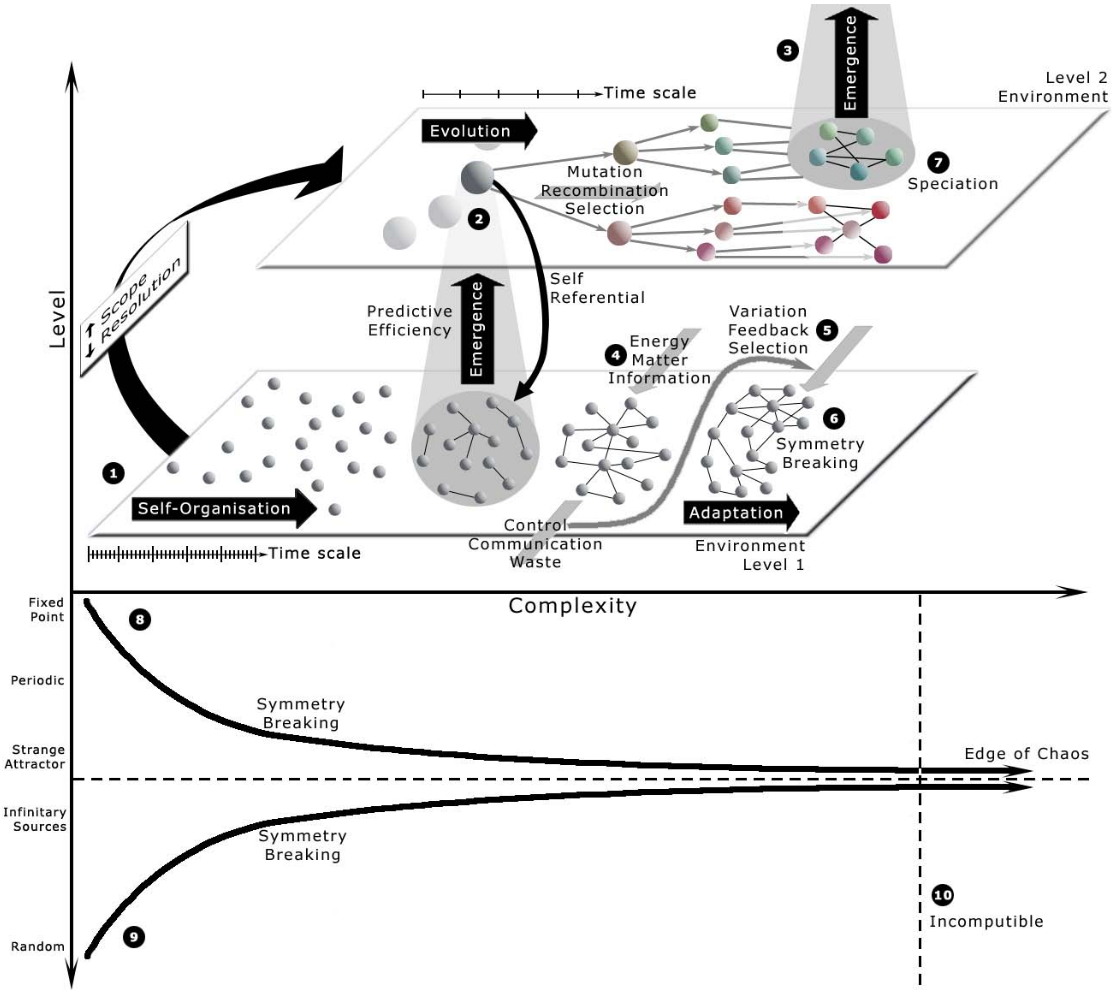

Figure 5.1: A systems view of complex systems concepts.

Figure 5.1 illustrates some relationships between the concepts introduced in this chapter. In particular, it shows two levels of an emergence hierarchy that are used to describe a complex system. The figure depicts dynamics that tend to increase complexity as arrows from left to right, and increases in the level of organisation as arrows from bottom to top. The concepts can be related in numerical order as follows. (1) demonstrates self-organisation, as components increase in organisation over time. As the components become more organised, interdependencies arise constraining the autonomy of the components, and at some point it is more efficient to describe tightly coupled components as an emergent whole, rather than a collection of unrelated elements. (2) depicts a lower resolution description of the whole. If the collective whole causally affects the behaviour of its components, it is self-referential. Note that Level 2 has a longer time scale. The scope at this level is also increased, such that the emergent whole is seen as one component in a wider population. As new generations descend with modification through mutation and/or recombination, natural selection operates on variants and the population evolves. (3) shows that interactions between members of a population can lead to the emergence of higher levels of organisation: in this case, a species is shown. (4) emphasises flows between the open system and the environment in the Level 1 description. Energy, matter and information enter the system, and control, communication and waste can flow back out to affect the environment. When the control provides feedback between the outputs and the inputs of the system in (5), its behaviour can be regulated. When the feedback contains variation in the interaction between the system and its environment, and is subject to a selection pressure, the system adapts. Positive feedback that reinforces variations at (6) results in symmetry breaking and/or phase transitions. (7) shows analogous symmetry breaking in Level 2 in the form of speciation.

Below the complexity axis, a complementary view of system complexity in terms of behaviour, rather than structure, is provided. Fixed point behaviour at (8) has low complexity, which increases for deterministic periodic and strange attractors. The bifurcation process is a form of symmetry breaking. Random behaviour at (9) also has low complexity, which increases as the system's components become more organised into processes with "infinitary sources" [95]: e.g. positive-entropy-rate variations on the Thue-Morse process and other stochastic analogues of various context-free languages. This class of behaviour ("infinitary sources") is depicted in (1). The asymptote between (8) and (9) is interpreted as the 'edge of chaos', where the complexity can grow without bound. Beyond some threshold of complexity at (10), the behaviour is incomputable: it cannot be simulated in finite time on a Universal Turing Machine.

For our discussion we chose an information-theoretical framework. There are four primary reasons for this choice:

1. It enables clear and consistent definitions and relationships between complexity, emergence and self-organisation in the physical world;
2. The same concepts can equally be applied to biology;
3. From a biological perspective, the basic ideas naturally extend to adaptation and evolution, which begins to address the question of why complexity and self-organisation are ubiquitous and apparently increasing in the biosphere; and
4. It provides a unified setting, within which the description of relevant information channels provides significant insights of practical utility.

As noted earlier, once the information channels are identified by designers of a physical system (or assigned to interactions between a bio-system and its environment by modellers), the rest is mostly a matter of computation. This computation can be decomposed into "diversity" and "equivocation", as demonstrated in the discussed examples.

Information theory provides a set of tools to carry out experiments, make predictions, and computationally solve real-world problems. Like all toolboxes, its application requires a set of assumptions regarding the processes and conditions regarding data collection to be satisfied. For example, information theory when used alone does not distinguish systems by spatial configuration. Also, it is by definition biased towards a view of Nature as an immense information processing device. Whether this view and these tools can be successfully applied to the large variety of problems complex systems aims to address is far from obvious. Our intent, at this stage, is simply to propose it as a framework for a less ambiguous discussion among practitioners from different disciplines. The suggested interpretations of the concepts may be at best temporary place-holders in an evolving discipline – hopefully, the improved communication which can arise from sharing a common language will lead to deeper understanding, which in turn will enable our proposals to be sharpened, rethought and continue to evolve.

Chapter 6

Emergence

Since its application to systems, emergence has been explained in terms of levels of observation. This approach has led to confusion, contradiction, incoherence and at times mysticism. When the idea of level is replaced by a framework of scope, resolution and state, this confusion is dissolved. It is shown that emergent properties are determined by the relationship between the scope of macrostate and microstate descriptions. This establishes a normative definition of emergent properties and emergence that makes sense of previous descriptive definitions of emergence. In particular, this framework sheds light on which classes of emergent properties are epistemic and which are ontological, and identifies fundamental limits to our ability to capture emergence in formal systems.

6.1 Introduction

The early development of emergence in the philosophical literature was in the context of the emergence of vitality from chemistry, and the emergence of minds from biology. This promoted the importance of understanding emergence, as a potential explanation not only of the relation between the general and special sciences, but also of the evolution of life, intelligence and complexity. Yet with its dual edge, the sword of the Emergentist philosophers carved out an overly ambitious research agenda, which set as its subject domain processes that are still largely impenetrable to science. Section 6.2 briefly reviews this history, a discourse that has largely obscured the fact that at heart, emergent properties are simply a difference between global and local structure.

The purpose of this chapter is to advance a new definition of emergent properties and emergence. Emergence is an essential pillar of every systems approach, and yet no precise, well defined account of emergence has achieved any level of consensus among systems researchers. The current surge of interest in complex systems – arguably the systems approach that seeks the closest integration with science – desperately lacks a clear understanding of emergence. For systems theory to be relevant to empirical experimentation, exact, testable concepts are necessary. It is not unreasonable to expect that achieving coherence on what it means for a system to have emergent properties, and when this counts as emergence, would lead to significant advances in systems research. I suggest that the current murkiness surrounding the central concepts in complex systems represents a serious impediment to progress. Clarifying emergence is an important step towards enhancing communication within the systems community. Even more importantly, it can improve communication with other fields of inquiry, enabling – among other possibilities – the application of exact systems concepts in science.

The approach taken in this study departs from the tradition of philosophical fascination in the emergence of life, consciousness and the universe. Undoubtedly, the most interesting exemplars of emergence are complex, highly evolved, and probably even self-organising. However, for emergence to be a useful and unambiguous distinction, firstly it must be isolable in a more basic form. Secondly, it must be understood in terms of well defined primitives. In Section 6.3, I argue that the conventional account in terms of levels and hierarchy does not meet the second criterion, so an alternative framework of scope, resolution and state is defined. In Section 6.4, an emergent property is defined, and simple examples show that novel emergent properties are coupled to scope. Section 6.5 defines emergence as the process whereby novel emergent properties are created, and examines the relationship between emergence and predictability. Section 6.6 uses the definition of an emergent property to outline a principled approach to determining the boundary of a system. I consider the practical limitations of the definitions in Section 6.7, and point towards some practical applications in Section 6.8.

6.1.1 Novel contributions

This chapter is based on:

[268] A. J. Ryan. Emergence is coupled to scope, not level. *Submitted to Complexity*, 2007.

My original contributions are:

1. A review of theories of emergence in philosophy and systems theory;
2. A precise definition of emergent properties;
3. Showing that the Möbius strip has novel emergent properties;
4. A precise definition of emergence;
5. Showing there exist fundamental limitations on generating emergence within formal systems;
6. A new method of deducing an objective system boundary from emergent properties;
7. Practical examples of emergent properties and emergence in different fields of science and engineering;
8. Discussion of the practical limitations for the application of emergence; and
9. The original insight that emergence is coupled to scope, not level.

6.2 A Short History of Emergence

The notion of an emergent effect was first coined in 1875 by the philosopher George Lewes[1] [203] to describe non-additive effects of causal interactions, to be contrasted with resultants. According to Lewes,

> Although each effect is the resultant of its components, one cannot always trace the steps of the process, so as to see in the product the mode of operation of each factor. In the latter case, I propose to call the effect an emergent. It arises out of the combined agencies, but in a form which does not display the agents in action ... Every resultant is either a sum or a difference of the co-operant forces; their sum, when their directions are the same – their difference when their directions are contrary. Further, every resultant is clearly traceable in its components, because these are homogeneous and commensurable ... It is otherwise with emergents, when, instead of adding measurable motion to measurable motion, or things of one kind to other individuals of their kind, there is a cooperation of things of unlike kinds ... The emergent is unlike its components in so far as these are incommensurable, and it cannot be reduced to their sum or their difference.

Several new theories of emergence appeared in the 1920s. Lewes' emergents were subtly refined by Alexander [6, 7], Broad [62] and Morgan [229–231], among others, to support a layered view of nature. The concept of emergence as a relation

[1] A precursor to emergence was the idea that the whole is more than the sum of its parts, which is usually attributed to Plato or Aristotle [14, 147]. Lewes was also influenced by Mill's [223] description of heteropathic effects, whose multiple simultaneous causes were not additive.

between simultaneous causes and their joint effect was translated to consider the upward causation of composition. Emergence now focused on properties rather than dynamical interactions, by considering the relationship between components and the whole they compose. This view on emergence was held in contrast to reductionist mechanism, the ideal that all apparently different kinds of matter are the same stuff, differing only in the number, arrangement and movement of their constituent components [62, p45]. Although properties of a complex whole were still simultaneously caused by the properties of components, novel system properties were said to emerge if they could not, even in theory, be deduced from complete knowledge of the properties of the components, either taken separately or in other combinations. Emergent properties were therefore irreducible, and represented barriers to mechanistic explanations. It was this conception of emergence that became the kernel of the mid twentieth century systems movement, as summarised by Checkland [81]:

> It is the concept of organized complexity which became the subject matter of the new discipline 'systems'; and the general model of organized complexity is that there exists a hierarchy of levels of organization, each more complex than the one below, a level being characterized by emergent properties which do not exist at the lower level. Indeed, more than the fact that they 'do not exist' at the lower level, emergent properties are *meaningless* in the language appropriate to the lower level.

Although Checkland suggests that the levels of hierarchy are ordered by complexity, he is in fact defining an emergence hierarchy[2], and there is no necessary condition on higher levels of organisation having greater complexity[3]. In any case, Checkland does explain the standard argument for emergence clearly. In this account, the macro language contains concepts that are meaningless at the micro level, in the way that it is meaningless to talk about flocking as a property of a single bird. It is important to emphasise that this does not claim one cannot map between sets of <u>microstate</u> descriptions to <u>macrostates</u>: it is the methodological assumption of supervenience that such a mapping is in principle possible[4]. Instead, it is the weaker assertion that when any component of a system is viewed *in isolation*, its microstate description cannot map to the associated emergent properties.

Since the systems movement adopted the conception of emergence as a relation between levels, explanations of emergence have diverged to include a remarkable number of contradictory positions. This includes a number of reductionist scientific explanations that erase the distinction between emergent properties and mech-

[2] An emergence hierarchy is a system view from a structural perspective made on the basis of the existence of emergent properties [72].

[3] Bar-Yam [25, p5, p746] provides several examples of emergent simplicity. Whether higher levels of an emergence hierarchy are necessarily more complex is dependent on whether the hierarchy is nested, and if the scope and resolution of description is the same at all levels. Checkland is by no means the only author to conflate an emergence hierarchy with increasing complexity; this has been common practice since the 1920s.

[4] The alternative position is a form of ontological pluralism, such as Cartesian substance dualism or organismic vitalism, which discourages scientific investigation. This approach declares at least one explanatory primitive (the mind or *élan vital*) which is by definition non-material and inaccessible to science.

anistic explanations, relegating emergence to the merely epiphenomenal[5]. Many other explanations tie emergence to evolution, complexity and/or self-organisation, presenting a singular unintelligible knot of concepts. Meanwhile, in contemporary philosophy a spectrum of conflicting positions, broadly either epistemological or ontological approaches to emergence, have been articulated with little headway made in either camp[6]. As the only commonality amongst the alternative positions is their failure to gain sufficient traction to generate consensus, their variety has only reinforced the status of emergence as an enigma.

6.3 Replacing Level With Scope and Resolution

The conventional explanation of emergence presented in the previous section is unsatisfactory. The use of an emergence hierarchy to account for emergent properties is alarmingly circular, given that the levels are defined by the existence of emergent properties[7]. In hierarchy theory, levels are most often considered to be epistemic, although seemingly only to avoid the burden of proof that falls on an ontological position. Many hierarchy theorists prefer to remain reality-agnostic [9]. Unsurprisingly, the inconclusive nature of levels means that explanations of emergence in terms of levels of description are unable to resolve its nature – is emergence a natural phenomenon or an artifact of the process of observation? To bypass this impediment, one needs to define emergence without invoking the concept of levels, which I argue can be accomplished using scope, resolution and state[8].

Scope is defined by a spatial boundary. Spatial is used in the broadest sense of the word to include conceptual and formal, as well as physical spaces, provided the system has a physical manifestation (spatial refers to the set of components, in contrast to temporal, which refers to the dynamics of those components). The scope of a system representation is the set of components within the boundary between the associated system and its environment. If an observer shifts from representing the system to representing a component, such that the component is now the system of interest, the scope of observation has narrowed. Conversely, when the scope is increased to include components that were previously part of the environment, the scope has broadened. There is also a temporal dimension to scope, which defines the set of moments of time over which the system is represented. $\mathcal{S}$ denotes scope, while $\mathcal{S}(x)$ and $\mathcal{S}(\tau)$ denote only the spatial and temporal dimensions of scope respectively.

Resolution is defined as the finest spatial distinction between two alternative system configurations. If a fine (high) and a coarse (low) resolution representation have the same scope, the fine resolution can distinguish a greater number of possibilities, n, and therefore each state contains more (Shannon) information,

[5]See the survey on The Laws of Emergence in [88, p 24].

[6]Matthews [216, p203] describes this impasse; also see [240] for a review.

[7]For a sounder (but still not explanatory) account of emergence in terms of levels see [70].

[8]Bar-Yam [30] recognises the importance of scope, scale and the microstate-macrostate relation in understanding emergence, and consequently [30] is closer to the following account of emergence than the sources covered in Section 6.2.

$H = -\sum_{i=1}^{n} p_i \log(p_i) = \log(n)$, assuming all states are equiprobable. A closely related concept is scale, which is a transformation by multiplication. The connection is that as a property is scaled up (multiplied) within a system, it can be detected at coarser resolutions. The distinction (which is rarely made) is that scale is independent of how the system is represented, whereas resolution is an attribute of the representation (scale is ontological, but resolution is epistemological). Once the resolution is set, this determines the 'size' of the components that comprise the system. There is also a temporal dimension to resolution, which defines the duration of a moment in time, where longer moments represent coarser (lower) resolutions. $\mathcal{R}$ denotes resolution, while $\mathcal{R}(x)$ and $\mathcal{R}(\tau)$ denote only the spatial and temporal dimensions of resolution respectively.

Figure 6.1 illustrates different temporal and spatial scopes and resolutions, reproduced from [266]. a) shows an *Elodea* leaf placed on a microscope slide. b) shows the same leaf at higher magnification: it has the same amount of data (the same number of pixels), but it has a narrower spatial scope and finer spatial resolution. c) contains four frames of a movie demonstrating streaming cytoplasm within the cell, and has a broader temporal scope than a) or b). The scope of c) is similar to a), but its spatial resolution is coarser.

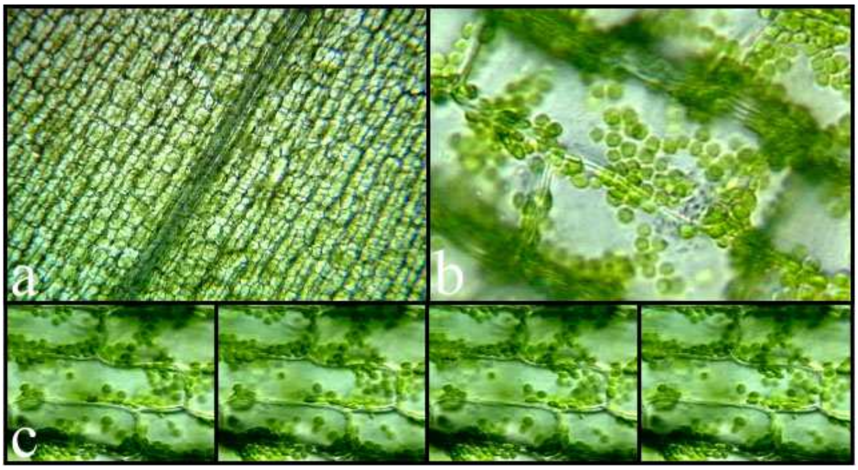

Figure 6.1: Example of different temporal and spatial scopes and resolutions.

The state of a system is the information that distinguishes between alternative system configurations up to some resolution at one moment in time. Macrostate M and microstate μ denote sets of states with two different resolutions and scopes, with the following macro-to-micro relations:

$$\mathcal{R}_{\mathrm{M}} \leq \mathcal{R}_{\mu} \tag{6.1}$$

$$\mathcal{S}_{\mathrm{M}} \geq \mathcal{S}_{\mu} \tag{6.2}$$

$$(\mathcal{R}_{\mathrm{M}}, \mathcal{S}_{\mathrm{M}}) \neq (\mathcal{R}_{\mu}, \mathcal{S}_{\mu}) \tag{6.3}$$

Intuitively, the macrostate has either a coarser resolution or a broader scope, or both. Let $\mathrm{M}' \in \mathscr{M}_{\mathrm{M}}$ and $\mu' \in \mathscr{M}_{\mu}$ denote the sets of $\mathrm{M}'|\mu$ and $\mu'|\mathrm{M}$ respectively satisfying Equations (6.1)-(6.3). Also note that M and μ represent sets of states if $\mathcal{S}(\tau) > 1$. This non-standard usage of the terms enables the representation of ensembles, and allows for emergent properties to be structured in time as well as space.

There exist other factors that influence the representation of a system by an observer. They include perspective (some information at a particular resolution is hidden eg. the state of internal organs to the naked eye) and interpretation (eg. optical illusions that have multiple valid interpretations). However, one does not need to invoke these factors to account for emergence, so for simplicity they are excluded.

6.4 Emergent Properties

Definition 6.1 (Emergent property)**.** *A property is emergent iff it is present in a macrostate and it is not present in the microstate.*

For this purpose, it is not necessary to impose any limitations on how the presence of emergent properties are inferred, provided the same methods are available in M and μ. The application of a set of methods designed to infer whether an emergent property is present or not constitutes a decision procedure. If the decision procedure returns a value of 1, informally the property has been detected. Ideally, an emergent property should be consistently present in an ensemble, to distinguish an emergent property from a statistically unlikely transient pattern (such as a recognisable image appearing in one frame of white noise). For a macrostate M, $P_{\mu}^{\mathrm{M}}(t) = \{\rho_1, \rho_2, ..., \rho_n\}$ is the set of emergent properties present in M and not present in μ at time t. If $\mathcal{S}_{\mathrm{M}}(\tau) > 1$, then t indicates the most recent moment in $\mathcal{S}_{\mathrm{M}}(\tau)$.

It follows from the definition that emergent properties must be the result of spatially or temporally extended structures, since otherwise it would be trivial to detect their presence in the microstate. By structure, I mean there is a pattern that relates the components, which implies redundancy, and therefore the description of the components is compressible. Structure means the components are 'organised' in the way Ashby [20] intended: communication (in some generalised sense) occurs between components to act like a constraint in the product space of possibilities. A corollary to this is that if the components are independent, they cannot give rise to emergent properties. Consequently, a Gaussian distribution is not organised, nor is it an emergent property of IID components. The law of large numbers is a statement about the loss of structure, not the emergence of new structure. Further, superpositionality, averaging and other linear operations cannot be the source of emergent properties. This is because a linear operator evaluates equally for any arrangement of the components. Because addition is commutative, linear operations capture a common feature of a set of components independent of their organisation, so the global structure is always exactly the sum of its parts. Lewe's original insight on emergents can now be restated: nonlinearity is a necessary condition for emergent properties.

So far, I have not specified which $\mathrm{M} \in \mathscr{M}_{\mathrm{M}}$ and $\mu \in \mathscr{M}_{\mu}$ were chosen as the macrostate and microstate in Definition 6.1. Given that $\exists \rho_i \in P_{\mu}^{\mathrm{M}}(t)$, we would like to know if ρ_i is a result of a change of resolution or scope. Both cases are now considered, by holding one of the two variables equal between the macrostate and the associated microstate.

6.4.1 Class I: Weak emergent properties

For the first case let $\mathcal{S}_{\mathrm{M}} = \mathcal{S}_{\mu}$, which by Equations (6.1) and (6.3) implies $\mathcal{R}_{\mathrm{M}} < \mathcal{R}_{\mu}$. Hence, $H(\mathrm{M}) < H(\mu)$. Let the surjective map[9] $\mathcal{C} : \mathscr{M}_{\mu} \rightarrow \mathscr{M}_{\mathrm{M}}$ give the coarsegrained macrostate corresponding to the microstate μ. By Definition 6.1, a decision procedure exists to detect ρ_i in M. But if ρ_i can be detected in $\mathrm{M} = \mathcal{C}(\mu)$, it can also be detected in μ by applying $\mathcal{C}$, followed by the decision procedure for M. Therefore, in this case, the presence of ρ_i in M implies the presence of ρ_i in μ.

The difficulty with determining the presence of emergent properties arises from finding the map $\mathcal{C}$ in the first place: it represents 'hidden structure', rather than 'novel structure'. This problem reduces to a combinatorial search problem for the mapping that reveals the relationship between the structure hidden in μ, and its more apparent representation in M. The worst scenario is if $\mathcal{C}$ is non-recursive, in which case the procedure for detection outlined above is incomputable. However, it is unknown whether physical processes exist that are capable of performing non-recursive mappings[10]. Aside from incomputability, the most challenging case is an incompressible iterative function, such that $\mathcal{C}$ involves a large but finite number of transformations. For example, consider a simulation of a discrete time dynamical system, where μ is a vector of the initial conditions and updating rules, and M is the binary terminal state of the simulation. $\mathcal{C}$ is incompressible if the most efficient way to infer M given μ is by running the simulation[11].

Properties in this case are classified as <u>weak emergent properties</u>, which is consistent with Bar-Yam's [30, p17] definition of "the relationship of microscopic and macroscopic views of a system that differ only in precision". A weak emergent property is epistemic, since once one has discovered the right mapping $\mathcal{C}$, by Defintion 6.1 it can no longer be considered emergent. Even in the extreme case of incomputability, it is a limitation in one's ability to detect the property that creates the appearance of it being emergent. In other words, one only believes a weak emergent property is not present in μ because of practical or fundamental limitations in the ability to detect and deduce the consequences of the structures that give rise to the emergent property in M. When practical limitations are the cause, a weak emergent property may appear to be emergent to one observer, but is not emergent to an observer with a deeper understanding of the microstate[12]. Within the assumptions and definitions of this study, if resolution is the only difference between a macrostate and microstate, no property of the macrostate can be genuinely emergent from the microstate.

[9] That such a function exists is the fundamental methodological assumption of science referred to in Section 6.2.

[10] See [87] for a review of the possibility of physical processes whose behaviour conforms to non-recursive mappings.

[11] This corresponds to Darley's [99] definition: "A true emergent phenomenon is one for which the optimal means of prediction is simulation". Holland's [154] book on emergence takes a very similar approach, without committing to a precise definition.

[12] In the literature, note that the label 'weak' is rarely used by proponents of this position. However, a purely epistemic conception of emergence is usually revealed by the assignment of emergence to the *relationship* between the observer and the system. A good example is Weinberg [324, p60], who states "We can always find cases in which a property will be 'emergent' to one observer and 'predictable' to another".

The class of weak emergent properties can be summarised by the following definition.

Definition 6.2 (Weak Emergent Property)**.** *A property is weakly emergent iff it is present in a macrostate but it is not apparent in the microstate, and this macrostate differs from the microstate only in resolution. A weak emergent property is a limitation of the observer, not a property of the system of interest.*

6.4.2 Class II: Novel emergent properties

For the second case let $\mathcal{R}_{\mathrm{M}} = \mathcal{R}_{\mu}$, which by Equations (6.2) and (6.3) implies $\mathcal{S}_{\mathrm{M}} > \mathcal{S}_{\mu}$. In this case, $H(\mathrm{M}) > H(\mu)$. It is useful to identify a minimal macrostate as a macrostate with the smallest possible scope that still exhibits ρ_i.

Definition 6.3 (Minimal Macrostate)**.** *A macrostate* M^* *is minimal with respect to an emergent property, if the emergent property is present in* M^**, and it is not present in any* μ *with the same resolution and narrower scope (ie. in any proper subset of the components of* M^**).*

M^* is not necessarily unique, since if two components are interchangeable, then there are two distinct M^* that can both satisfy Definition 6.3. Three simple examples will show that M^* can be well defined.

Firstly, consider a Möbius strip, which is a one sided, one edged, non-orientable 'surface with boundary'. One can think of the Möbius strip as being comprised of a singly twisted loop of triangles, such as the tiling depicted in Figure 6.2. It can be shown that any compact differentiable manifold allows a triangulation. $\mathcal{R}_{\mathrm{M}^*}$ is determined by the number of triangles used, and $\mathcal{S}_{\mathrm{M}^*}$ equals the set of triangles. If one considers any proper subset by removing at least one triangle, the resulting surface or surfaces are two sided, orientable and have more than one edge. Formally, the Euler characteristic χ is 0 for a Möbius strip, but equals the number of disjoint simplicial (triangular) complexes in μ. This is equal to or greater than 1, so M^* is not topologically equivalent to any μ. Therefore, the properties associated with the Möbius strip are emergent properties of M^* that do not exist for narrower scopes. Further, as χ is a topological invariant, it does not depend on the resolution of the triangulation. Hence, the emergent property is coupled to the scope of M^*, irrespective of which particular surfaces are defined as the components of M^*.

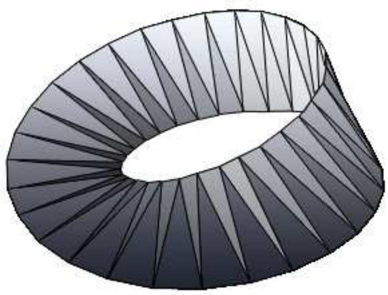

Figure 6.2: A Möbius strip can be triangulated to show it has novel emergent properties.

Secondly, a perfect secret sharing scheme divides some data D into n pieces $D_1, ..., D_n$ such that:

1. Knowledge of any k or more D_i pieces makes D easily computable; and
2. Knowledge of any $k-1$ or fewer D_i pieces leaves D completely undetermined (in the sense that all its possible values are equally likely).

An efficient perfect secret sharing scheme is presented in [286] based on polynomial interpolation, where a random polynomial is chosen of degree $k-1$ such that $q(x) = a_0 + a_1 x + ... + a_{k-1}x^{k-1}$ has $a_0 = D$ and $a_i \in [0, p)$ are integer and bounded by a large prime p. Then the keys $D_i \in \mathbb{Z}$ are generated by $D_i = q(i) \mathrm{mod}\ p,\ i = 1, .., n$. Interpolation retrieves a unique value for $a_0 = D$ if at least k values of $q(i)$ are known, but when $k-1$ values are revealed to an opponent, all polynomials associated with possible secrets D' are equally likely – the mutual information $I(D; D') = 0$. Thus when $\mathcal{R}_{\mathrm{M}} = \mathcal{R}_\mu = \min(|D_i - D_{i'}|) = 1$ (ie. one can distinguish between each possible $D_i \in [0, p)$) and $\mathcal{S}_{\mathrm{M}^*} = k$, then D is present in M^* and is not present in any μ. By construction, D is an emergent property that is coupled to any M^* with a scope of k.

The third example presents an instance where the emergent property depends on temporal rather than spatial scope. Conceptually, there is no difference between structure extended in space or time, except that communication can only move forwards in time, meaning that structure can only constrain possibilities that are within the future light cone of the first component of the structure in the temporal scope. Consider a process governed by the deterministic periodic discrete time iterative function

$$f(t + Pn) = f(t) \quad \forall t \in \mathbb{Z}, \tag{6.4}$$

where $P \in \mathbb{Z}^+$ is the period and $n \in \mathbb{N}$, such that Pn is a multiple of P. The property of this sequence is translational symmetry, which is present in M^* when $(\mathcal{R}_{\mathrm{M}^*}(t), \mathcal{S}_{\mathrm{M}^*}(t)) = (1, P+1)$, and is not present in any μ with a $\mathcal{S}_\mu(t) \leq P$. This example is trivial, but the extension of the idea of temporal emergent properties to general discrete dynamical systems includes far more interesting structures.

This class of emergent property arises from structure that is extended over the scope of the system, which I refer to as novel emergent properties. There is a difference between local and global structure in any system that exhibits emergent novelty. This explains why emergent novelty cannot be understood or predicted by an observer whose scope is limited to only one component of a system. When the resolution of the macrostate equals the resolution of the microstate (or the inverse mapping from the macrostate to a set of microstates is known and well defined), emergent novelty is at least in part ontological. One cannot say it is fully ontological, since some *a priori* concepts are used in this framework to structure the analysis. However, the minimal macrostate has an objective property that is independent of variations in the epistemic status of an observer.

Novel emergent properties are not just nonlinear, they are nonseparable. This is because if they were separable, then their presence could be established without requiring a scope at least as great as the minimal macrostate.

Definition 6.4 (Novel Emergent Property)**.** *A property is a novel emergent property iff it is present in a macrostate but it is not present in any microstate, where* *the microstates differ from the macrostate only in scope.*

One subclass of emergent novelty that has been discussed as a separate phenomenon in the literature is 'emergent behaviour', which is a property of the system that is only exhibited in certain environments. An example of emergent behaviour is the interaction between a lock and a key[13]. The key is said to have an emergent behaviour, since it opens any door containing a complementary lock, and this is not present in the microstate description of the key *in isolation*. This property is only present in the macrostate when the spatial scope of the system is expanded to include both the lock and the key. The emergent behaviour can then be explained as complementary spatially extended structure between two system components. From this analysis, I conclude that emergent behaviours are the result of mistakenly attributing a novel emergent property of a system to one of its components. While it is often convenient to keep the idealised system boundary fixed and talk of emergent behaviours, at the same time one should be clear that the scope of the emergent property extends between the system and certain contexts.

The classification of emergent properties has not considered the case that $\mathcal{R}_{\mathrm{M}} < \mathcal{R}_{\mu}$ and $\mathcal{S}_{\mathrm{M}} > \mathcal{S}_{\mu}$. This case is harder to analyse, since it is not possible to say whether $H(\mathrm{M})$ is greater or less than $H(\mu)$. Fortunately, because one can rule out resolution as a source of emergent properties, the only important factor is scope. Therefore, simply by coarsegraining μ so that $\mathcal{R}_{\mathrm{M}} = \mathcal{R}_{\mu}$, this is reduced to Class II.

6.5 Emergence

So far I have analysed emergent properties without saying how they arise. This is the process of emergence.

Definition 6.5 (Emergence)**.** *Emergence is the process whereby the assembly, breakdown or restructuring of a system results in one or more novel emergent properties.*

Assembly and breakdown are the dual processes of adding and removing interactions between system components that change the cardinality of the set of components in the system, while restructuring changes interactions between components without changing the cardinality. Some researchers claim that only self-assembling and self-restructuring processes can emerge. For example, Holland [154] argues that emergence must be the product of self-organisation, not centralised control. This is another example of attempting to tie together separate concepts that are useful only if they have distinct meanings. It usually leads to circular definitions (emergence is self-organising; self-organisation is a process that gives rise to emergence), or greedy reductions (emergence is nothing-but self-organisation).

[13] Following Bar-Yam [30], I use this example for ease of comparison with what Bar-Yam calls 'type 3 strong emergence' or 'environmental emergence'. Additionally, a popular example of emergence in the literature is the smell of ammonia, which is said to emerge from the odourless nitrogen and hydrogen components. What is not usually explained is that the smell of ammonia is a property of the relationship between the gas and human olfactory receptors, and is therefore an example of environmental emergence. According to 'shape' theories of olfaction, the interaction between ammonia molecules and the human receptor system is not dissimilar to a lock and key.

When an emergent property is reinforced by positive feedback, its scale is increased. However, it is important to emphasise that merely scaling an emergent property is not emergence. The common use of the word does not make this distinction. For example, consider the following observation on avian bird flu. "Despite the widespread emergence of H5N1 influenza viruses in poultry in many countries in Asia in early 2004, there were no outbreaks of H5N1 influenza in poultry or humans in Hong Kong during this time." [211] In this case, "widespread emergence" is synonymous with "growth". It refers to scaling of the population infected by H5N1, rather than the assembly of the initial mutation of the virus. Scaling is an important process, since if an emergent property is not reinforced it cannot perpetuate or have a significant impact (consider a mutation that does not replicate). However, the technical definition of emergence only applies to the initial process of assembly.

Having made this distinction, one can now make a useful observation on the connection between centralised control and emergence. Centralised control is characterised by a lack of autonomy in the system's components, except for the controller. The controller can assemble (or breakdown or restructure) the other components of the system, which may result in a spatiotemporally extended property. If this is performed by following some template or blueprint, one can ask whether the emergent property is present in the template. If so, then the property emerged within the controller, then was scaled up (which is not emergence) as it was realised across the system's components. If not, one can say the property emerged in the system, even though it was assembled under central control from a template. If there is no template, then the emergence also occurs in the system, not the controller. Error-free, context-insensitive asexual replication can only scale existing emergent properties, but the introduction of mutation, crossover, retroviruses or 'nurture' (ie. sensitivity to information from the environment) can lead to emergence. In summary, emergence can occur through centralised control, provided the emergent property is not already present in the controller.

Emergence is defined above as a process, which means it is temporally extended. That is, emergence is not a property of a system at any point in time, it is a relationship between system properties at two different moments in time. Let M have emergent properties $P_{\mathrm{M}}(t) = \{\rho_1, \rho_2, ..., \rho_n\}$ at time t. At some later time t', the system's emergent properties are $P_{\mathrm{M}}(t') = \{\rho_1, \rho_2, ..., \rho_{n-r}, \rho_{n+1}, ..., \rho_{n+s}\}$, with $r, s \in \mathbb{N}$. If $\max\{r, s\} > 0$, then at least one new emergent property is present, or a previous emergent property no longer exists in M. In either case, between t and t', M exhibits emergence[14].

As a thought experiment, one can set $\mathcal{S}_\mu$ to be unbounded in space and include all prior moments in time. In this case, M with the same resolution and spatial scope contains just one new moment in time. Now, if M contains a novel emergent property $\hat{p}_i$, then $\hat{p}_i$ has never existed before. For M* such that $\hat{p}_i \in P_{\mathrm{M}^*}$, $\mathcal{S}_{\mathrm{M}^*}$ contains a corresponding structure whose specific configuration has likewise not occurred previously. I coin the term 'naissance emergence' to refer to the original emergence of a novel emergent property $\hat{p}_i$. Naissance emergence is the source of

[14]Note that every ρ_i has a logical complement, 'absence of ρ_i'. Therefore the disappearance of ρ_i is logically equivalent to the appearance of its complement, and vice versa. This is why it is not possible to separate the roles of assembly and breakdown in emergence.

novelty, and is an important distinction for a discussion of the relationship between emergence and predictability.

The problem for scientists aspiring to *predict* naissance emergence is that, by definition, $\hat{p}_i$ is not present until it is within temporal scope ie. until it has already occurred! Of course, a scientist may have a theory about what properties may pertain for a configuration that has not existed. But from Section 6.4, a theory where the proposed properties are linear combinations of properties of the components in other subsets or configurations cannot give rise to a novel emergent property. Also, from Section 6.4.2, recall that novel emergent properties are nonseparable. Therefore, any theory that claims to predict naissance emergence must extrapolate $\hat{p}_i$ from a nonlinear and nonseparable combination of previously observed properties. But if the extrapolation is nonlinear, it is not unique. Therefore, our scientist must have multiple theories for $\hat{p}_i$, all of which are possible. There is no logical way to choose between the candidate theories, so a choice of $\hat{p}_i$ can only be justified by empirical experiment. But by conducting the experiment, $\hat{p}_i$ is now within temporal scope. Consequently, $\hat{p}_i$ cannot be predicted with certainty until it has already occurred.

This implies that formal systems, including mathematical models and computer simulations, are incapable of reproducing naissance emergence. This does not mean that once naissance emergence has occurred that one cannot alter one's models to include $\hat{p}_i$ and associate it with some M*. It just means that this cannot be done *a priori*, because empirical access is necessary to select between the possible properties of completely new configurations. Naissance emergence is an ontological concept, since in light of the preceding discussion it cannot be epistemic.

6.6 Rethinking System Boundaries

Interestingly, the view of emergent properties developed in this study presents an alternative way of defining the boundary of a system. It is rare that the process of system definition is treated explicitly. However, it is suggested that the following process is typical. Firstly, the system boundary is chosen to separate the system from its environment where the interactions are weakest, which sets the scope. Weak interactions are targeted because open systems will always have flows of inputs and outputs across the boundary, but if these flows are weak compared to internal interactions, they can be either ignored or aggregated, and a systems analysis should accurately capture first order features of the system.

Secondly, once the scope of the system is set, deciding the resolution is usually straightforward. If the scope is the biosphere, it is currently infeasible to model at the resolution of individual molecules. If the scope is an individual molecule, then the resolution will need to be significantly finer if one wants to say anything useful about the system. Limitations on the available cognitive and/or information processing resources provides an upper bound on the practical resolution for observing a system of any given scope, and since the upper bound is Pareto dominant with respect to information, observers could be expected to be near the Pareto dominant values (ignoring extremely small scopes). Thus, limitations on information processing provides an approximately linear inversely proportional relationship between scope and resolution. Just like a camera's zoom lens, varying the scope automatically adjusts the resolution, and automatic processes are subconscious and hardwired, rather than conscious, deliberate and justified.

Thirdly, now that the scope and resolution are known, the system is composed of a finite number of components. Emergent properties belong to the system if they do not occur in the absence of the system, and are not properties of the components taken separately or in other combinations.

A number of issues arise from this kind of approach to defining a system. The most obvious problem is because it is subconscious, intuitive and unstated, it is not subject to criticism or debate. The definition of the system is axiomatic, and despite its crucial role in the success or failure of all subsequent analysis, it is placed beyond question, or rather slipped in beneath questioning. In addition, the open nature of most systems of interest means specifying a unique boundary is problematic. At what point does the flow of matter, energy and information stop being part of the environment and start being part of the system? Finally, I showed in Section 6.4.2 that this approach leads to the idea of emergent behaviours, whereby emergent properties are assigned to the system, when the system is only one component of the structure extending between the system and its environment that gives rise to the emergent property.

Definition 6.4.2 enables an alternative approach for identifying the boundaries of a system. Firstly, a system is defined by a set of properties $\{\rho_1, \rho_2, ..., \rho_n\}$ that characterise and identify that system. Secondly, for each property i, the minimal macrostate M_i^* is identified, which associates that property with a particular scope, $\mathcal{S}_{\mathrm{M}_i^*}$. Thirdly, the system boundary is defined as the set union of the scope for each property, $\bigcup_{i=1}^{n} \mathcal{S}_{\mathrm{M}_i^*}$. Finally, the resolution must be at least as fine as the highest resolution minimal macrostate. By starting from a set of emergent properties, the process is explicit and justified; flows from the environment are included in the system boundary only when they are a necessary component of a system property; and every property must belong to a subset of the system's components. The ontological nature of novel emergent properties means the system boundaries derived from them are not arbitrary, but reflect features of the system that are independent of the observer.

It may be argued that in practice, it is never possible to list every property of a system. This is not a major limitation due to the nature of the union operator. When a subset of the set of all system properties is considered, $\{\rho_1, \rho_2, ..., \rho_k\}$, with $k < n$, the derived system boundary represents a lower bound on the actual system boundary. This is easy to see, since $\bigcup_{i=1}^{k} \mathcal{S}_{\mathrm{M}_i^*} \subseteq \bigcup_{i=1}^{n} \mathcal{S}_{\mathrm{M}_i^*}$. Further, as new properties are added to the subset, the derived system boundary must converge from below to the actual system boundary. Using this procedure, one can approach a representation of a system's boundary, whose only dependence on the observer is deciding on the set of properties to be associated with the system.

6.7 Practical Limitations

The analysis above has helped to clarify the role resolution and scope play in emergent properties of substantial systems with a unique, well defined microstate. Mathematical examples are useful because truth is accessible within the rules of the axiomatic system. It is possible to analytically show properties of microstates and macrostates, avoiding problems such as the theory-ladenness of observation that arise when properties must be detected empirically. Unfortunately, a number of such issues limit the ability to decisively show the presence of emergent properties and emergence in the real world. One of the most difficult aspects of identifying emergent properties in natural examples is choosing the resolution for the macrostate, which determines what are considered to be components. If a novel emergent property is present with respect to one set of components, but not for another way of defining the components, is the property really emergent? If there exists *any* way of defining the components, such that the emergent property is present in a microstate with narrower scope, then the novel emergent property does not belong to the macrostate. The property is still emergent, but it has been attributed to the wrong scope, because of the choice of resolution. In general the decision procedure cannot check every possible resolution, so in practice applying Definition 6.4.2 could overestimate the scope of the minimal macrostate.

Microstates and macrostates are defined in Section 6.3 to cater for an ensemble perspective. Although this has not been required in the examples above, except to capture temporal structure, it will often be required for physical systems. When considering sufficiently fine microstates (either quantum or semi-classical), observations of a system over time cannot be performed on the same microstate, but rather on the ensemble of states. In this situation, a single microstate is not physically observable and therefore is not a physically meaningful concept [30]. This means one needs to take an ensemble perspective [30], which complicates the process of observation by making it statistical, but it is still entirely compatible with the framework developed in this study.

To quote Zadeh [336], "[m]ore often than not, the classes of objects encountered in the real physical world do not have precisely defined criteria of membership". If a property is either statistical or fuzzy, there will not be a discontinuous boundary between emergent and non-emergent. For example, a self-avoiding random walk has the property of almost completely unstructured movement when the scope of the microstate is one move, and the property of considerably structured, statistically self-similar movement when the scope of the macrostate is a large number of moves. The emergent property of statistical self-similarity is satisfied with greater confidence as the temporal scope of the macrostate broadens. In this case, the scope associated with the novel emergent property will be somewhat arbitrary. However, one can at least specify bounds on the associated scope, such that below the lower bound the novel emergent property does not exist, while it exists with an arbitrary degree of confidence above the upper bound on scope.

There exist many systems that are not studied according to the distribution of their physical substance, including formal (mathematical) systems and social systems. In these systems, a convenient property of physical systems is absent. In physical systems, entropy is well defined by the quantum difference given by Planck's constant h [25, p13]. This means there exists only a finite number of distinct possibilities. Even though formal and social systems must both ultimately have physical instantiations, they do not have obvious bounds (analogous to Planck's constant) on possibilities. For instance, although the number of distinct thoughts a human mind will have in its lifetime is finite, one apparently cannot specify in advance any finite set containing every possible thought, nor determine the finest possible distinction between two thoughts the mind is capable of making. In mathematical systems, the bounds on scope and resolution are even less obvious (real numbers in general contain infinite information), which is why mathematics is sometimes described as the study of all possible worlds. If the distinct possibilities are not bounded, then resolution may not be a meaningful concept, and the microstate may contain infinite information. In order to apply this framework to non-physical systems, spaces must first be approximated with a finite product space of possibilities, so that the microstate and resolution are both well defined.

6.8 Practical Applications

The preceding mathematical examples are useful because of their simplicity and precision. This section will not be precise or conclusive, but rather suggestive of real world examples of emergence. The aim is to provide a few hooks for some of the many disciplines that investigate emergent phenomena.

Simple machines such as pulleys and levers provide a mechanical advantage by decreasing the amount of force required to do a given amount of work. In isolation, simple machines increase the scale of effect of the energy source, which by Section 6.5 is not emergence, or transform energy between forms. However, when a collection of simple machines are assembled to form a compound machine, it is possible that the compound machine has an emergent property that no proper subset of the parts and the energy source exhibit. Usually, these emergent properties are thought of as either the function of the compound machine, or unintended consequences. A particularly vivid example of emergent functionality is given by Rube Goldberg machines, named after the American cartoonist who drew improbable machines for performing simple tasks. The board game Mousetrap, and the computer game The Incredible Machine, are both instantiations of Rube Goldberg machines where no proper subset of the system components can achieve the functionality of the complete system. Almost all engineered systems – clocks, radios, computers, and aeroplanes – are designed for specific, predictable emergent properties. However, note that to provide robust functionality, most engineered systems contain redundancy, which means the system contains more components than the minimal macrostate. Also note that if the emergent function is a behaviour, then by Section 6.4.2 the emergent property is formally a property of the larger system in which the engineered system is used.

In chemistry, a catalyst decreases the activation energy of a chemical reaction. An autocatalytic set is defined as a reaction system (M, R) of molecules M and reactions R, such that all the catalysts for all its reactions R are in M. If no proper subset of (M, R) is an autocatalytic set, then the reaction system has the emergent property of catalytic closure. Some researchers, such as Kauffman [175, p329], have linked certain forms of catalytic closure with the ability to generate heritable variation, and consequently to evolve under the pressure of natural selection. Although the speculative claims of artificial chemistry have not yet been empirically demonstrated, autocatalysis is an obvious candidate for the emergence of novelty in chemistry.

In biology, many synergies have been studied that may be examples of emergence. The link between synergy and emergence has been made by Corning [88], who defines synergy as "the combined (cooperative) effects that are produced by two or more particles, elements, parts, or organisms – effects that are not otherwise attainable". Once again, a cautionary remark obtains that greater efficiency through synergy is just scaling, not emergence. One synergy that does appear to be due to emergent properties is obligate endosymbiosis, such as the relationship between the *Olavius algarvensis* Oligochaete – a gutless marine worm – and the chemoautotrophic bacteria that lives inside it. Under certain conditions (such as the absence of an external source of reduced sulphur compounds), neither organism can survive in isolation. However, a syntrophic sulphur cycle recycles oxidised and reduced sulphur between the symbionts, which is believed to have enabled *O. algarvensis* to colonise new habitats and extend their geographic distribution [113].

Another biological example demonstrates how emergent properties can be either spatially or temporally extended. Cyanobacteria are a remarkably diverse group of prokaryotes. Different species that engage in the chemically incompatible processes of nitrogen fixing and photosynthesis have evolved different solutions to work around this obstacle. *Anabaena* spatially separates the processes in separate heterocysts and passes the products between cells using filaments [288]. In contrast, *Synechococcus* temporally separates the processes, performing photosynthesis during the day and nitrogen fixation at night [111]. Yet another species, *Trichodesmium* both spatially and temporally separates the processes [45]. All three species have the same emergent property, which since it cannot be a property of a single component, must be distributed in space, time, or both space and time.

In economics, games such as the tragedy of the commons [145] and prisoner's dilemma [259] capture a mathematical representation of the difference between local and global structure. In these games, pursuit of a local maximum in the microstate (one player's payoff) prevents the players from maximising the macrostate (the total payoff to all players). The structure of the games and greedy rational behaviour combine to ensure a sub-optimal Nash equilibrium will prevail, even when a solution exists where all players could have received greater payoffs. This kind of outcome results because the scope of consideration of greedy rational players is too narrow. If the negative externalities of the actions of individual players are incorporated into the payoff (which broadens the scope of what each player pays attention to), then the structure of the game changes and the equilibrium will no longer be dominated by other combinations of strategies. The game is now factored, meaning any increase in the payoff to one player does not decrease the total payoff to all players [334]. The point at which incorporating the cost of negative externalities into the payoff matrix results in a factored game is the scope associated with this emergent property. The applications of game theory are of course much wider than just economics. There are many games with a similar difference between local and global structure, such as Braess' Paradox [58], where adding extra capacity to a network can reduce global performance, and Parrondo's paradox [146], where playing two losing games can be a winning strategy overall. In Parrondo's paradox, scope has a different meaning (playing each game in isolation or in conjunction), but the emergent property is still coupled to scope.

6.9 Summary

Due to its central position in systems approaches, a redefinition of emergence has significant implications for systems research. The definitions in this study do not directly contradict the common view that emergent properties at one level are meaningless at the level below. However, they do provide a deeper understanding that represents a substantial refinement of the common conception. A simple explanation of emergent properties is given in terms of scope, which forms the basis of a normative definition of emergent properties and subsequently for emergence. That is, rather than just describing what emergent properties are like, my definition prescribes the conditions whereby a property should be formally considered to be emergent.

Definition 6.1 is so general it is trivial. Given almost any macrostate, one can choose a microstate with a sufficiently small scope that the macrostate has an emergent property relative to that microstate. If almost every macrostate has emergent properties, the definition is meaningless. This is why the idea of the minimal macrostate M^* is crucial. M^* allows emergent properties to be coupled to a specific scope. Definition 6.4.2 is more specific, because it only counts emergent properties when they are not a property of *any* microstate with smaller scope. Consequently, only a small fraction of macrostates have novel emergent properties.

A number of phenomena have previously been lumped under the banner of emergence. In this study, I found that the concept is currently too broad. Weak emergent properties must be excluded from emergence: the resolution of observation, or the language of description has no bearing on whether a property is emergent. Emergent behaviour, or environmental emergence, must be reassessed as a novel emergent property of a system with larger scope. A clear distinction was made between emergent properties and emergence, which shows that simply scaling an emergent property cannot be considered emergence.

Lewes thought of emergents as the converse of resultants, while Broad recast emergentism in opposition to mechanism. In this study, an emergent property is the converse of a local property. This is consistent with Lewes, insofar as resultants represent linear combinations of existing localised components. Nonlinearity is a necessary but not sufficient condition for an emergent property. In contrast, if a property is non-local, it is spatially or temporally extended, and necessarily emergent, so this opposition is more revealing. The relationship with Broad's antonym is less clear, since it depends on how strictly mechanism is interpreted. Under the purest interpretation of mechanism, local properties should tell us everything there is to know about a system, and there is no potential for naissance emergence. However, pure mechanism is not really a serious metaphysical position, and contrasting it with emergentism does little to reveal what emergent properties are. The view of emergence in this study seems to be most compatible with non-reductive physicalism, although it is not my intent to advocate a particular ontology. In summary, Definition 6.1 not only enables us to say what emergent properties are, it also allows us to better say what they are not.

As long as there exist possible configurations of the universe that have not yet occurred, one must be realistic about our ability to predict future dynamics on the basis of formal systems, when the dynamics may be influenced by emergence. This insight is the most profound implication of emergence, yet also the most difficult aspect of emergence to demonstrate constructively, by virtue of its absence in formal systems.

The alternative to level and hierarchy – scope, resolution and state – offers a generic framework for analysing systems. Because the primitives are well-defined for any physical system, they should have broad applicability in systems research. An obvious direction for further research is to supplement these primitives with other concepts to provide a more powerful formalism.

Given that science has always valued depth of knowledge over breadth, it is not surprising that a scientific understanding of emergence has not been forthcoming, when emergent properties are precisely the properties that cannot be understood with additional depth. The existence of emergent properties provides legitimation for broader multidisciplinary systems approaches that complement specialised scientific disciplines. However, this insight is not articulated in discussions of emergence based on levels, because in general a level is an indeterminate mixture between relationships of scope and resolution. By revealing the coupling between emergence and scope, it is hoped that the dialogue on emergence can achieve coherence.

The above analysis of emergence concludes the development of the theoretical framework for this thesis. Having addressed questions on the nature of representation, systems and multidisciplinary research, and then provided novel formulations of complex systems concepts including emergence, self-organisation and adaptation, it is now possible to investigate the application of multidisciplinary conceptual analysis to real world problems. Part III of this thesis introduces two problems that are amenable to a complex systems approach. They are real world problems with real stakeholders, so it is important to stress that they have not been 'reverse engineered' to exemplify the preceding conceptual framework. Although convenient toy models could have been developed to showcase the conceptual framework, it was decided that a more flexible application of the framework to contemporary defence problems would better demonstrate the potential impact of multidisciplinary conceptual analysis.

The first case study tackles the problem of conventional force structures conducting operations against an asymmetric insurgency. This problem was raised due to current operations in the Middle East, and both drew on and contributed to Future Land Warfare's concepts for operating in this environment – Complex Warfighting [181] and Adaptive Campaigning [318]. The second case study responds to a request from Capability Development Group to provide suggestions for a more flexible and agile capability development process, compared with the current process as documented in the Defence Capability Development Manual [22].

Part III

Application

Chapter 7

Asymmetric Warfare

This chapter applies multidisciplinary conceptual analysis to the problem of force design for asymmetric operations. Conventional military forces are organised to generate large scale effects against similarly structured adversaries. Asymmetric warfare is a 'game' between a conventional military force and a weaker adversary that is unable to match the scale of effects of the conventional force. In asymmetric warfare, an insurgents' strategy can be understood using a multi-scale perspective: by generating and exploiting fine scale complexity, insurgents prevent the conventional force from acting at the scale they are designed for. Insights from complex systems show how future force structures can be designed to adapt to environmental complexity at multiple scales – including complexity generated by asymmetric adversaries – to achieve full spectrum dominance.

7.1 Introduction

The cold war was a story about two giant bears, Uncle Sam and Mother Russia, armed to the teeth and locked in the dynamic stability of mutually assured destruction. The nuclear arms race saw both bears grow stronger and more dangerous. The only non-catastrophic resolution could be voluntary withdrawal by one side, which occurred when Mother Russia backed down, exhausted and weak from keeping pace with Uncle Sam. Now the undisputed power in the woods, Uncle Sam felt lost at first without the focus that a life-threatening adversary demands. Searching for animals that might one day become powerful bears, Uncle Sam encountered many mystical animals: dragons, elephant gods and a reptilian Godzilla, although none seemed interested in confronting the dominant bear on his terms.

So Uncle Sam went into hibernation, until one September several sharp stings rudely awaked him. Enraged, the bear found the small bodies of the bees that had released their venomous barbed stingers in his flesh[1].Uncle Sam tracked the honey trail to a hive in the mountains, which he smashed with his giant paws. Unsatisfied, the bear continued over the mountains and into the desert where he knew of another beehive, rumoured to have even nastier stings. This hive too was quickly smashed, but Uncle Sam was in for a surprise. For this species of desert bee was more aggressive and easily agitated. Every time they stung the intruder, an alarm pheromone triggered more bees to attack. The swarming enemy was small, mobile and dissolved into the desert under direct attack, which rendered the bear's keen eyesight and large muscles ineffective and impotent. Crushing individual bees could not destroy the swarm, and even when the bear had limited success, a Goliath triumph over David could do little to win the hearts and minds of the neighbouring animals. Truth be told, the locals rather thought the bear was stirring up a hornet's nest.

This story illustrates just a few of the issues associated with the transition from the cold war era to the so-called Global War on Terror. It is a problem that is relevant to every nation state in every region. Terrorism is asymmetric, enduring, ancient, unencapsulated and continually co-evolving: it is the kind of problem that has most stubbornly resisted conventional scientific reduction, the kind of problem that has helped motivate the rise of systems thinking.

Section 7.2 provides insights from complex systems theory, explaining in terms of organisation and scale why the bear is ineffective against the bees, and introduces adaptation as a possible response. The conjunction of systems engineering and complex systems (beginning to be recognised as a separate field as complex systems engineering) is developed in Section 7.3 to provide a new approach to developing capabilities for complex environments. Section 7.4 explains the attrition warfare paradigm that conventional forces are organised to fight, which is contrasted with asymmetric warfare in Section 7.5. Adaptive responses to asymmetric warfare are discussed in Section 7.6.

[1] I thank Martin Burke for suggesting this metaphor for asymmetric warfare. Swarms of bees are also discussed as a metaphor for effects-based operations in [295].

7.1.1 Novel contributions

This chapter is based on:

[267] A. J. Ryan. About the bears and the bees: Adaptive approaches to asymmetric warfare. *Interjournal*, 2006.

My original contributions are:

1. Development of the relationship between scale and complexity in asymmetric warfare;
2. Identification of adaptive responses to asymmetric warfare, involving changes in force structure and capability development.

7.2 Two Complex Systems Theories

Two complex systems theories are useful for understanding asymmetric warfare: multiscale variety and adaptation. The **law of multiscale variety** [31, 32] is an extension of Ashby's law of requisite variety [18] (see Section 3.3.2). Assume that a system has N parts that must be coordinated to respond to external contexts, and the scale of the response is given by the number of parts that participate in the coordinated response. Secondly, assume that under (complete) coordination, the variety of the coordinated parts equals the variety of a single part. Then, coordination increases the scale of response, but decreases its variety: there is a tradeoff between large scale behaviour and fine scale complexity. In this chapter complexity refers to the (log of the) number of possible configurations, which is a measure of variety based on Shannon information. The generalised law of multiscale variety states that at every scale the variety necessary to meet the tasks, at that scale, must be larger for the system than the task requirements [32].

Adaptation is a generic model for learning successful interaction strategies between a system and a complex and potentially non-stationary environment (see Section 5.7). The environment is treated as a black box, and stimulus response interactions provide feedback that modifies an internal model or representation of the environment, which affects the probability of the system taking future actions. For the case that the system is a single agent with a fixed set of available actions, the environmental feedback is a single real valued reward plus the observed state at each time step, and the internal model is an estimate of the future value of each state, this model of adaptation reduces to reinforcement learning. However, an adaptive system may also be comprised of a number of agents acting and sensing the environment in parallel, in which case the representation is distributed. The distributed representation may not be consistent or complete, but can be thought of as an ecology of cooperating and competing models, each partially representing some aspects of the environment. The three essential functions for an adaptive mechanism are generating variety, observing feedback from interactions with the environment, and selection to reinforce some interactions and inhibit others.

7.3 Complex Systems Engineering

The capability development process (See Chapter 8) is responsible for acquiring all new major capabilities in defence. Consequently, it has a significant impact on force structure. In order to change the way coalition forces are structured for asymmetric operations, it is necessary to examine force design and acquisition. In contrast to the classical systems engineering approach to this problem, here a complex systems engineering approach is introduced.

Complex systems engineering when taken literally is a contradiction in terms. If conventional approaches to modelling do not capture the dynamics of a complex system, one cannot expect to predict the effects of design choices on behaviour, an essential precondition for engineering. Therefore, either the systems one can engineer will never actually be behaviourally complex (and certainly no more complex than the designer), or else a new understanding of what it means to design and engineer a system is required. Since there appears to be no bound to the complexity of problems people can imagine engineering solutions to, the latter course of action seems inevitable. Complex systems engineering will be examined in greater detail in Chapter 8, but some aspects that are relevant to the problem of asymmetric warfare will be highlighted.

In order to increase the complexity of engineering designs, the first step is to surrender the security blanket of complete understanding and control for systems design. Real limits on certainty and predictability of system behaviour should be acknowledged, and also that complete test and evaluation is not feasible [28, Thm A]. Rather than planning a design centrally, the design occurs distributed across a team with managerial independence, relying more on market forces and end user feedback than the ability of the lead systems engineer to understand and predict the behaviour of the whole system.

Each designer would be responsible for developing a functional building block for the system, capable of operating independently. The way in which different functions interface would not be pre-specified. Interfaces would be negotiated locally as the building blocks are developed, creating large numbers of possible interaction networks. Whereas traditional systems engineering takes a glueware approach to integration that tends to minimise the number of possible interaction patterns [236], harnessing self-organising mechanisms to produce global structure provides an important source of variety. Even when the components of a self-organising system are fixed, the multiple possible arrangements of components provides the flexibility to respond to unexpected environments and exploit unplanned functionality.

Traditional systems engineering assumes fixed epochs, where one system is replaced by a new system that is essentially designed from scratch. However, complex problems will not have fixed solutions, so the design, develop, operate, and dispose life cycle is artificial. Rather than one-off replacement costs, complex systems engineering projects could be given a continual flow of funding that incrementally changes the system. The system is always operational, and the distinction between legacy and current systems is dissolved. Because the interaction patterns are not fixed, as new components are added the system will self-organise to include them.

Some of the new components will directly compete with legacy components for links. As the new components demonstrate better performance, their interactions will be reinforced, while old components that fall into disuse can be removed. The coexistence of generations in the one system provides redundancy [27], which alleviates the need for exhaustive test and evaluation and allows greater risk-taking and innovation when designing new components.

7.4 Attrition Warfare

Conventional military forces are organised to generate large scale effects against similarly structured adversaries. Industrial age mass production allowed nation states to re-equip and reorganise rapidly following defeats in battle that in the Napoleonic era would have been decisive [295]. Therefore, industrial age armies had to be prepared to fight wars of attrition, where the primary aim is the physical destruction of the adversary's armed forces and its support base. Because of this, the attrition 'game' is dominated by the physical resources or mass each player can bring to bear. Each side competes to produce effects on a larger scale in order to overwhelm their opponent's defences. The Lanchester differential equations are the classical model for attrition warfare because they capture the importance of mass in achieving the physical destruction of an opponent.

The original form of the Lanchester equation for directed fire warfare [192] is:

$$\frac{dm_B}{dt} = -rm_B \tag{7.1}$$

$$\frac{dm_R}{dt} = -bm_R, \tag{7.2}$$

where m_B, m_R are the number of force units in the Blue and Red sides, their derivatives represent the attrition rate, and b, r are the effective firing rate of a single force unit for the Blue and Red forces respectively. If the initial size of Blue and Red is M_B and M_R, the solution to the differential equations can be rearranged to give Lanchester's square law:

$$\frac{b}{r}\frac{M_B^2}{M_R^2}\frac{m_B(t)^2}{M_B^2} - \frac{m_R(t)^2}{M_R^2} = \frac{b}{r}\frac{M_B^2}{M_R^2} - 1 \tag{7.3}$$

Intuitively, Equation (7.3) means that increases in initial mass M_B will produce quadratic improvements in battlefield superiority for the Blue force compared to improvements in efficiency b.

Of course, the Lanchester model is a representation that captures the dynamics of combat only in the limit of the simplest environment. This is because the model assumes that every combatant is within detection and engagement range, forces are homogeneous, fire is uniformly distributed over surviving units, and factors such as manoeuvre, logistics, and command and control have no significant effect on casualties. As an environment becomes more complex, mass becomes increasingly irrelevant since the complexity limits the size of effects that can be massed against any target. Consequently, spatial arrangement becomes increasingly important, which suggests that a systems approach may be more appropriate than aggregate models.

So how is a conventional force organised? Consider two organisational drivers for conventional military structures. Firstly, there has been a trend throughout the history of warfare towards greater lethality at all levels of warfare in order to mass a large scale effect in the shortest possible time frame. As is expected from the multiscale law of requisite variety, this emphasises greater coordination of the activities of units at all levels. "Synchronisation" enables the exploitation of synergies and compensation for weaknesses of different units within combined arms teams, and at a higher level of organisation by joint warfighting. Centralised control, detailed planning and collective training exercises are the dominant mechanisms for producing synchronised effects. A continual focus on synchronisation to produce large scale effects has encouraged specialisation. Specialisation can increase efficiency by avoiding redundancy, but it also locks in dependencies between units, reducing the ability of units to operate in isolation at finer scales. In order to focus effects, conventional forces are organised to be spatially oriented towards the forward edge of the battle area, with most protection for the front-line manoeuvre units, less protection for stand-off indirect fire units and little protection for the rear logistics supply chain.

Secondly, with increasing lethality comes increasing responsibility and the potential for the misuse of force to adversely affect the strategic goals of conflict. This has reinforced the use of centralised control and codified procedures to limit flexibility and prevent strategically detrimental application of force. Highly centralised hierarchical control reliably amplifies the scale at which the commander can control battlespace effects, at the expense of the fine scale variety (complexity) that the conventional force can cope with.

7.5 Asymmetric Warfare

In contrast to attrition warfare, asymmetric warfare is a 'game' between a conventional military force and a weaker adversary that is unable to match the scale of effects of the conventional force. To compensate, the weaker adversary must believe their will to accept the costs of conflict is greater than their opponent. What began as Operation Iraqi Freedom, and is now referred to by the Pentagon as the Long War, is a paradigmatic example. The US-led Multinational force is opposed by an Iraqi insurgency composed of approximately 14 guerilla organisations, each with distinct aims and relatively independent operations. Although the insurgents have limited material means, their religious and ideological motivations provide strong will. A history of aversion to casualties in democracies during conflicts, especially when the conflict is peripheral to the nation's core interests, provides a basis for the insurgents' belief that the Multinational force's strategic will can be weakened if domestic support for a continued presence in Iraq is sufficiently eroded. At the tactical level, suicide bombings represent an asymmetry in will, where suicide bombers have sufficient will to utilise methods outside those available to the Multinational force.

For insurgents to exploit their asymmetries, they must also negate the asymmetries that favour the conventional force. In particular, they must avoid direct, large scale confrontation against the better equipped, trained and synchronised conventional force. This can be understood using a multi-scale perspective: by generating and exploiting fine scale complexity, insurgents prevent the conventional force from acting at the scale they are organised for – large scale but limited complexity environments.

By dispersing into largely independent cells, insurgents can limit the amount of damage any single attack from the conventional force can inflict. This significantly reduces the threat of retaliation from acting as a deterrent, since the insurgents have negligible physical resources exposed to retaliatory attack [295]. Insurgents that do not wear uniforms and blend into a civilian population cannot be readily identified or targeted until they attack, in a situation of their choice. There is no longer a forward edge of the battle line, meaning softer support units are vulnerable. The number of possible locations, times and direction of attack increases significantly compared to attrition warfare, increasing fine scale complexity. The heightened potential for collateral damage from mixing with civilian populations dramatically increases the task complexity for a conventional force that must minimise the deaths of innocent civilians for any hope of strategic victory.

By moving into complex terrain, such as urban or high density vegetation, the Multinational force's most expensive sensors (such as satellites and radar) designed to detect large scale movements are ineffective due to the fine scale of physical activity, which provides a very low signal to noise ratio. In this situation, insurgents control the tempo and intensity of the conflict, which enables them to exploit niches in the fuzzy spaces near artificial boundaries, such as the traditional conceptual dichotomies of war/peace, combatant/non-combatant, state/non-state actors, and tactical/strategic operations, adding further to complexity. The ability to mass synchronised battlespace effects is of little use in such a complex situation.

7.6 Adaptive Responses to Asymmetric Warfare

Clearly the conventional force requires a different organisation to respond to asymmetric powers, while still maintaining the ability to generate large scale effects when required. The force must be able to generate sufficient variety in effects at every scale, from peace-keeping and peace enforcement crises to high intensity warfighting, to achieve full spectrum dominance. Because the biggest capability gap is currently in asymmetric warfare, I will focus on this context.

Within the Australian Army, units form Platoons or Troops, each with a specialist function, such as to detect, respond or sustain. The Platoons in a Battlegroup can then be combined to form Company level combat teams, and integration at lower levels is uncommon. In contrast, the Australian Complex Warfighting concept [181] outlines a new type of organisation to cope with increased complexity. Smaller, austere, semi-autonomous teams with modular organisation are envisaged. These teams use swarming tactics and devolved situation awareness to operate as self-reliant teams that aggregate to achieve larger scale effects through local coordination rather than central control. Teams would not be as specialised as current units, since each team must be largely self-reliant for logistics, sensing, decision-making and responding to threats. The difference between a logistics and reconnaissance team for complex warfighting would be a matter of emphasis, since both would have a base capability for mobility, survivability, detection and response. Complex warfighting tasks such as asymmetric warfare require a shift in conventional force structure towards Special Forces structures.

Each of these changes are consistent with complex systems insights. However, the new organisation presents additional challenges. Whereas a centralised system promotes standardisation of equipment and process, semi-autonomous teams will deliberately promote variety. As well as reflecting differences in local context, the variety will exist because some teams will discover successful strategies that are unknown to other teams. Therefore, it is necessary to promote the spread of successful strategies between autonomous teams to improve overall force effectiveness. However, if successful variations are adopted too readily, the reduction in variation between teams will diminish the force's ability to adapt. There exists a tradeoff between being adapted (specialised) to the current environment, and adaptability for future contexts. Numerous examples of this tradeoff between exploitation and exploration were provided in Chapter 5.

Another challenge associated with semi-autonomous teams is ensuring that the goals of the autonomous teams are aligned with force level goals. Abu Ghraib is an example where the pursuit of local goals (extracting intelligence) produced extremely damaging strategic effects. The impact of socio-cultural issues on strategic success is clear in this example. The interplay between increasingly ubiquitous mass media and populations that value surprise in news (which Shannon's theory accounts for) act to reinforce the perception of anomalies, and can also reinforce the public response. For armed forces, the more independent teams become, the more likely at least one team will develop a culture that reinforces strategically counterproductive behaviour. The role of higher level headquarters that manage semi-autonomous teams will be to clearly communicate and then police the boundaries of acceptable behaviour, within which autonomous teams have freedom to innovate.

This kind of force structure places very different demands on the capability development process. Whereas a conventional force demands standardisation so that effects can be synchronised and the output of units is predictable, complex warfighting teams will be heterogeneous and have varied demands for materiel. An industrial age approach to capability development tailored to large scale production of standardised materiel is unsuited to meeting the fine scale but complex demands of the new force structure. In contrast, the complex systems engineering model introduced in Section 7.3 enables a much more responsive development process. Individual autonomous teams could test new ways to sense and respond to asymmetric threats, enabling them to adapt at the secondary level. From the capability development perspective, semi-autonomous teams provide an ideal entry point and realistic test bed for new systems, and allow the coexistence of legacy and experimental systems discussed in Section 7.3. If the system provides a significant benefit, demand for it will quickly spread through the network of teams. If a new system has a negative effect on one team's performance, at least it will have a negligible effect on overall force performance due to the autonomy of the teams.

In order to be effective, the adaptive response must occur at all levels. At the strategic level, the use of complex warfighting teams to deny targeting success is only one available strategy in a space that includes economic, political and information operations. In this space, the effects of taking different sets of actions is typically far less certain and first, second and third order adaptive cycles can play a crucial role in identifying and exploiting useful sets of actions.

7.7 Conclusion

Asymmetric warfare presents a challenge of increased fine scale complexity. Current force structures are monolithic bears designed for large scale effects, and must be reorganised by devolving autonomy and increasing independence to provide sufficient fine scale variety. Once variety is available, the best way to cope with complexity is using adaptation, which can improve system performance over time and track changes in the environment. Complex systems theory has the potential to improve the way military capability is acquired, organised and managed, enabling adaptive responses to asymmetric warfare.

Chapter 8

Capability Development

This chapter develops the second case study for multidisciplinary conceptual analysis, investigating options for a more agile capability development process. The extant capability development process is a top-down planning, project-based acquisition methodology underpinned by traditional systems engineering. In this chapter we present an alternative process for capability development based on complex systems engineering and an entrepreneurial model for project management, which enables self-organisation with respect to capability requirements. Using complex systems theory we contend that this provides a basis for agile and flexible capability development, overcoming systemic issues with the current process. In particular, self-organising capability development is likely to result in better reuse, interoperability and integration as well as rapid satisfaction of new requirements.

8.1 Introduction

If the Australian Defence Force (ADF) is viewed as a system, capability development is one of the key points of influence where the decisions made can dramatically alter the composition and ability to fight for the future force. At a recent workshop on Complex Adaptive Systems in Defence, syndicate groups were presented with the following symptoms by Capability Development Group that indicate the capability development process may not be delivering capability in line with the expectations of the government, the media and the ADF itself [242]:

> **Background.** The [Australian Defence Organisation] has had a history of struggling to optimise strategies for force development. A number of systems, models and processes have been adopted and discarded. While agility and adaptability are recognised as desirable characteristics of future force structures they could also be considered desirable characteristics of a capability development model or process.
>
> **Symptoms.** The problem is primarily manifested in continuing with capability decisions that time has rendered irrelevant, ineffective or sub-optimal. The failure to identify the impacts of one capability decision on others produces similar effects.

The words that Capability Development Group (CDG) uses to express their symptoms are insightful into the underlying dynamics of the problems they face. The classic cause of the first symptom is developing stable future plans within a non-stationary context. The planning process is so inflexible that the course of action is invariant in the face of change. By the time the plans have been implemented, changes to the context render decisions made at an earlier date inappropriate. The sunk cost effect [15, 228] of developing a plan and committing to acquiring capabilities that will be less effective than previously envisaged can exacerbate the consequences if decision-makers are tempted to "bend reality to fit their plan".

The second symptom is typically caused when a complex problem is segmented into modular sub-problems. Placing artificial boundaries within the system has the advantage of constraining the degrees of freedom for each sub-problem and thus reducing its complexity. In capability development this is achieved by giving program managers nearly independent projects, each with independent measures of success. This strategy can be successful for simple problems, but as the problem becomes more complex (ie. the number of interdependent projects increases), the strategy does not scale, and the interfaces quickly become more complex than the individual projects [33].

The side effect of discretising capability into projects is that when they are delivered to the end user, the assumptions that were made in one project may affect another project, reducing the effectiveness of the total capability. Because program managers are goal-driven towards and rewarded for delivering against the requirements within their project, there is little incentive to divert resources from internal deconflicting to external deconflicting of requirements and assumptions. It is little surprise that under these circumstances the artificial boundaries produce stovepipes that are reinforced over time, reducing the ability of the capability development system to deliver holistic capabilities integrated across multiple projects.

Since a "number of systems, models and processes have been adopted and discarded", this suggests that one or more core assumptions are shared amongst previous approaches to develop capability. After uncovering some implicit assumptions previous approaches have made that are not appropriate for the problem of capability development, we will suggest an alternative methodology to overcome systemic causes behind the symptoms that have been observed in the current capability development process. The purpose of this chapter is not to present a water-tight alternative to the current process. Rather, it explores how else the capability development process could be. This is just the first step in the search for a way of organising decision-makers, such that the process delivers capabilities that are more closely aligned with the expectations of Government, citizens and soldiers.

8.1.1 Novel contributions

This chapter is based on:

[270] A. J. Ryan and D. O. Norman. Agile and Flexible Capability Development. In *Proceedings of the Land Warfare Conference*, Brisbane, Australia, 2006.

Our original contributions are:

1. A new way of assessing the fitness of capabilities;
2. A new approach to designing complex, tailored capabilities;
3. A deductive approach to comparing force structures; and
4. An entrepreneurial model for project management.

My role in this work was:

1. After an initial two days of jointly whiteboarding the scope, research questions and outline of the paper, I performed all of this work. The framework of questions we asked Capability Development Group to structure their problem was developed jointly between Anne-Marie Grisogono and myself.

8.2 The Current Capability Development Process

The current capability development process is described in detail in the Defence Capability Development Manual [22]. Capability is defined as "the power to achieve a desired operational effect in a nominated environment, within a specified time, and to sustain that effect for a designated period. Capability is delivered by systems that incorporate people, organisation, doctrine, collective training, platforms, materiel, facilities, in-service support, and command and management".

This emphasises that capability is a systemic property that arises from the interactions between the interdependent Fundamental Inputs to Capability (FIC), and links it to achieving a specific operational effect. Although the current capability development process is centred on the acquisition of materiel, namely major capital equipment investments of $20 million or more (or with individual items of $1 million or more), its measure of success is the based on the capability generated by the set of FICs operating as an organised whole.

Capability systems have a life cycle that includes phases to identify needs, develop requirements, acquire solutions, maintain in-service, and dispose of or redeploy the system. The Chief, Capability Development Group (CCDG) shares the responsibility of the needs phase with the Deputy Secretary, Strategy. CCDG has the primary responsibility for the requirements phase, and is a stakeholder in the acquisition and in-service phases. Consequently, we will limit the scope of this chapter to the needs and requirements phases of capability development.

The aim of capability planning is to "develop and maintain the most cost efficient and operationally effective mix of capabilities for achieving the Australian Government's strategic objectives". There are seven principles that guide the current process. They are: 1) the process is top-down, so that the value of different capabilities is assessed by and traceable to their contribution to the government's strategic priorities; 2) a bottom-up perspective examines how current and future planned building blocks of capability can best contribute operationally; 3) value for money, which considers the whole of life cost against the benefit of a capability; 4) a long range view is necessary due to the lag before capabilities enter service and the duration in-service once a commitment is made; 5) flexibility, which is the capacity of the capability development system to respond to changes in context and short notice requirements; 6) be informed by new operational concepts and how they may most efficiently utilise materiel to generate capability; and 7) evaluate capabilities with respect to adversary capabilities rather than specific threats, since capabilities are more enduring than threats.

It is interesting that there exist direct conflicts within the first five principles, which means capability development decisions will inevitably involve negotiation, tradeoffs and balancing of the development principles, rather than simultaneous optimisation. If the top-down and bottom-up perspectives come to opposing conclusions, this must be resolved before a project's priority can be agreed upon. A project that delivers long-term capability must be balanced against more immediate requirements in order to be flexible. And value for money involves balancing operational effect with the opportunity cost of other operational effects that the ADF must forgo in order to commit to a capability.

The capability development process is illustrated in Figure 8.1, which is reproduced from the DCDM [22]. The process begins at the top left corner with the development of high-level strategic documents including the White Paper, Defence Updates and the Budget. These documents set the broad capability priorities from Government specifying the strategic interests and tasks for the ADF. In the second stage, each of the broad strategic priorities is developed into military strategic objectives, effects and options in the Australia's Military Strategy document. They are related: the effects are chosen to meet the objectives, while the response options are selected to meet the effects required. The context for analysing different response options is provided by the Australian Illustrative Planning Scenarios (AIPS), joint and single service experimentation and operational concepts. The Defence Planning Guidance developed in the top right corner organises defence contingencies into priority bands based on Government guidance.

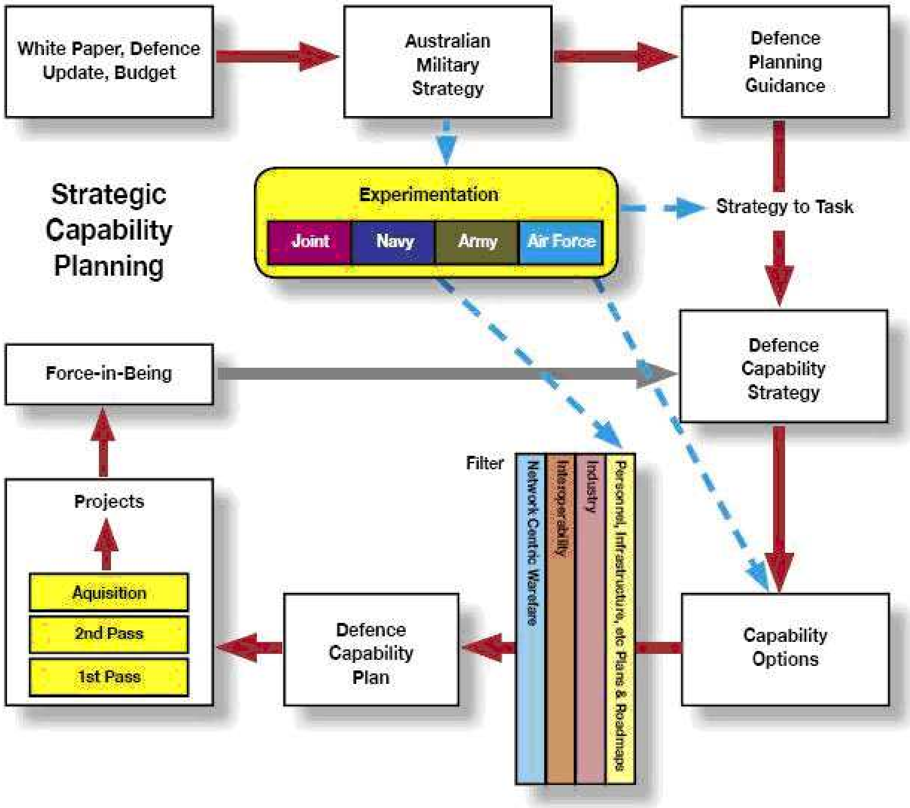

Figure 8.1: Overview of the Capability Development Process, after [22].

The Defence Capability Strategy identifies broad military capabilities required in each contingency, and evaluates how the requirements can be achieved at the minimum cost through sustaining or retiring current capabilities and acquiring new capabilities. This is an important stage since it compares the prioritised strategic requirements developed in previous stages with the Force In Being (FIB). The difference between the FIB as well as projects that are already in progress, and the requirements is a set of capability gaps. The ability to fill a capability gap is the primary justification for the entry of a new capability option into the Defence Capability Plan (DCP).

Any organisation or group in Defence can suggest capability options (these are expressed as an operational effect or a broadly defined equipment solution) in the form of an Initial Capability Definition Statement (ICDS). All ICDSs are reviewed by the Options Review Committee, known as the troika because it has three members: Chief, Capability Development Group; Head, Capability Systems; and First Assistant Secretary, Capability, Investment and Resources. The troika filters the list of ICDSs, resulting in an endorsed list that is further developed into a Capability Definition Statement (CDS). Once the CDS has been completed it returns to the troika. If it is endorsed the CDS is given a project number and becomes part of the Defence Capability Plan (DCP), which also means it is accepted by Government as an official need. The DCP is a rolling ten year plan of all unapproved Major Capital Investment projects, and is the critical link between strategic guidance and detailed capability planning. The completion of the DCP marks the end of the needs phase of the capability development process.

The requirements phase consists of progressing projects through the two pass approval system and results in a tender quality proposed solution agreed by government. Projects are developed for first pass committee approval by producing the Capability Options Document, which lists at least three broad options that fulfil the requirement. Projects that are approved on second pass have formal agreement from Government to fund the acquisition of a specific capability system with a well-defined budget and schedule. The details involved in first and second pass approval are now reproduced from the DCDM (sections 3.10, 3.11):

> First pass approval is, in effect, approval in principle to proceed with more detailed analysis and costing of broad capability proposals. More specifically, first pass approval provides government approval of:
>
> a. the broad functions and performance of the proposed capability;
>
> b. the proposed year of decision (YOD), Initial Operational Capability (IOC) and Full Operational Capability (FOC). The IOC is the point in time when a capability system is deemed by the Capability Manager as ready to be deployed in an operational role. The FOC is the point in time when there is Operational Acceptance or Operational Release of the whole capability, jointly ratified by the Sponsor and Capability Manager;
>
> c. the set of feasible generic options to be explored in more detail;
>
> d. the broad costing for the capability options, including FIC aspects;
>
> e. the timings of the development of the option(s);
>
> f. studies or other activities to support the development of options for the second pass, and funding to pursue those studies/activities;
>
> g. industry engagement to develop a business case for second pass approval; and
>
> h. the funding needed for Defence to undertake the detailed analysis of the approved options, including any test and evaluation (T&E) required.
>
> Second pass approval is formal approval by government of a specific capability solution to an identified capability development need. Second pass approval provides government approval of:
>
> a. a preferred specific capability solution chosen from the narrow band of options approved at the first pass approval stage;
>
> b. the specific functions and performance of the proposed capability;

c. the planned Year Of Decision (YOD), Initial Operational Capability (IOC) and Full Operational Capability (FOC);

d. budgetary provision for acquisition and operation of the capability solution, including FIC aspects;

e. the strategy for acquiring the capability and bringing it into service; and

f. T&E concepts and funding.

The intellectual basis for the process is the field of systems engineering, which was covered in Section 3.3.4. Traditional systems engineering is a highly effective methodology when certain assumptions are met. The requirements for the system should be known in advance, they should not be subject to change within the lifetime of the project, and resources across the whole process should be centrally controlled and able to be reallocated [236]. Unfortunately, none of these assumptions hold. The capability development process results in force structures that are unique for each operation, exist in a changing context, and are not developed under fully centralised control. It has been demonstrated that individual projects – let alone the whole capability development process – exist beyond the boundary conditions for the successful application of traditional systems engineering [236]. However, guidance for this domain is not entirely absent. Complex systems engineering, while still in its infancy, promises to address increasingly complex real-world problems using complex systems techniques [59].

8.2.1 Requirements

Based on the presentation by Capability Development Group [242], an improved capability development process would deliver more agile capabilities that are able to quickly assemble into forces to meet new mission operational requirements, with reduced friction and increased synchronisation and efficiency. Rather than take this as our starting point, we firstly examine why this may be desirable.

The specific operations the future force will undertake are unknown. Although the primary dimensions that characterise future threat environments are known, we cannot rule out the emergence of a new dimension. An example of this occurring is the use of nuclear weapons in World War II, which produced a new dimension of Weapons of Mass Destruction for both capability and the threat environment. Consequently, it is reasonable to assume that future threats will not only be non-stationary, but may develop in dimensions of which we are currently unaware. This has implications for measuring risk, since the theory of probability does not apply - it is the difference between an uncertain and an unknowable threat.

In addition, while Western defence forces have become increasingly organised to mass large scale effects, the current asymmetric threats are smaller scale but high complexity threats (see Chapter 7). In order to generate capabilities that are structured for complex warfighting, the capability development process will need to shift from large scale capability development to complex capability development. This is a key distinction that will be examined in Section 8.3.2.

It is worth explicitly identifying the link between threats and the capability development process. The type of threat the future force faces will shape the force structures that can successfully achieve their mission. The force structures that are required to succeed against a given threat dictate the requirements for materiel that the force structures consist of. The materiel requirements in turn shape the process of developing capability, since the process must be responsive to changes in the threat environment leading to demand for new materiel. Figure 8.2 summarises this causal relation.

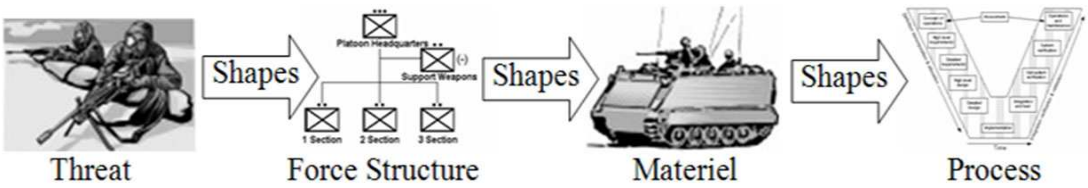

Figure 8.2: The influence of the nature of threat on the capability development process.

Figure 8.2 provides the business case for improving the agility of the capability development process. As long as the threats the ADF face can change and evolve, the capability development process must be flexible and agile if it is to supply effective capabilities in this environment.

The detailed requirements for a more flexible and agile process are now developed. CDG identified a number of difficulties with the current process [242]:

> **Issues.** Some of the issues include:
>
> - Monitoring and identifying changes to original assumptions behind decision making and feeding this back in to effect change,
> - Map, synergise and deconflict interactive processes such as intelligence, strategy, capability development, shaping and influencing, concept development and experimentation both functionally and across Defence outputs,
> - Map interactions between needs, requirements, acquisition, in service and disposal phases, and
> - Enhancing agility and adaptability characteristics of the system-optimising decision making and decision makers.

The first issue shows an implicit desire to develop and maintain a central list of changes to assumptions to identify points of impact. Such a centralised repository is likely to become outdated and add unnecessary overheads to the projects by continually requesting data in the struggle to remain relevant. This also assumes that changes to assumptions are understandable and identifiable as a coherent, consistent set. If this is not the case then centralised problem solving reduces to trial and error [171]. The requirement we can extract from this issue is the ability of the process to adapt to changes in assumptions, whether these are monitored and understood centrally or not.

The second issue indicates that the capability process cannot be separated from other functions within the ADF such as intelligence and concept development. There is no clear boundary separating capability development from the rest of the ADF, and since the rest of the ADF consists of autonomous actors who are not under the direct control of CDG, decisions within the capability development process must be negotiated with the other autonomous actors in order to achieve a successful outcome for the ADF. This must be explicitly acknowledged in the capability development process in order to deliver successful outcomes. In particular, it must be reflected in the way the success of individuals is measured.

Once the assumption of a centralised mapping mechanism is removed from the third issue, we require that the capability development process manage interactions across the functional boundaries within the capability development system. This is because if the functions coordinate their activities they can explore a greater region of outcome space, which we expect will include regions that deliver more capability to the end user. Therefore we need guidelines for the interactions between functional areas that enable the capability development system to find and exploit desirable regions of the outcome space.

The fourth issue implies that the current process is rigid, and asserts that a more adaptive and agile process will lead to more optimal decision making. While we agree that adaptation and agility are desirable attributes of the process, it should be noted that no matter how adaptive a process is, it is likely to satisfice rather than optimise within a complex system [11]. An excessive focus on optimisation can lead to improvements in efficiency, while the more important issue of effectiveness is ignored.

CDG have identified features of the process that lead to complexity [242]:

> **Complexity.** The complexity comes from:
>
> - Lag between a control and its effect - decisions on force structure taken now may not be implemented for 2-15 years. How to ensure the relevance of decisions made in very different contexts,
> - The interaction between elements of capability and understanding how changes to one element impact upon others,
> - The interaction between a number of autonomous actors and processes which are not constant and also change with time, and
> - The impacts of inputs from outside actors (Government, Industry) that may not be consistent with any logic/process used within Defence.

The first and forth items may add to problem difficulty, although they do not add to complexity in the technical sense of the word. However, the second and third items are rooted in complexity. The implications of each of these factors for the capability development process are now considered.

From control theory, we know that controlling even a linear time-invariant system using a static output feedback (or showing such a controller does not exist) is an open problem [300]. The long time lag identified in the first item above can lead to common control problems such as overshoot and oscillation. We have shown earlier the system is not time-invariant, which makes the control problem analytically far more difficult. This reinforces the requirement to produce a process that is adaptable rather than optimised.

The second item identifies the fundamental cause of complexity, which is interdependence between parts of a system. In any operation, the products of the capability development process must interact in order to generate capability. Because the current process acquires new equipment out of context in stovepiped projects, a project could meet its internal specifications, but have unintended consequences on other projects when they are required to interact. In order to avoid this outcome, we need mechanisms for revealing any conflicts with existing or planned capabilities, and negotiating decisions between a project and the capability owners they impact upon.

The third item requires that the autonomous actors be explicitly accounted for in the process, and importantly must recognise that actors change due to the posting cycle, while external agencies may also reorganise and change their procedures and processes. These complexities contribute to uncertainty, eroding the effectiveness of long-term planning and analysis.

The final item notes that the Government controls the direction, the requirements and the budget for defence, all of which may change without warning. The layer of Government control above the Australian Defence Force cannot be within the scope of any changes to the capability development process, and therefore the process must be designed to work within the Government layer. In addition, the ADF can only acquire equipment that industry can supply. The capability development process will enhance its long term effectiveness if it can foster a strong local industry, which may place pressure to provide a degree of certainty in decision-making and thereby limit short term adaptability.

CDG have provided the following list of metrics [242]:

> **Metrics.** An improvement would be observed as reduced friction and increased synchronization and efficiency of capability development processes. This might be evidenced by:
>
> - Reduction in time taken in requirements and acquisition phases of capability, development cycle and corresponding freeing up of man hours,
> - Greater confidence in process,
> - Reduction in delivery of capability that is missing some vital supporting component,
> - Reduction in delivery of capability where initial need has changed, and
> - Reduction in delivery of capability where requirements have been superseded.

In summary, a new process should aspire to be simpler and streamlined while delivering a capability the end user is happier with. Changes to the process should result in greater adaptivity, robustness to change, and better management of internal and external interactions across system boundaries and between autonomous actors.

8.3 New Perspectives

Different perspectives on capability development lead to different ways of thinking about problems and solutions. In this section we characterise the current paradigm, then develop new perspectives for measuring fitness, acquiring capability, assessing the gap between needs and capability, and controlling capability development decisions. Our characterisation of the current perspective is broad and non-specific, because it is intended to apply equally to previous similar approaches to capability development.

8.3.1 What is fitness for a capability development process?

The measure of fitness we use to compare different capability development processes will shape the characteristics of the process we select. The current process can be better understood once the benchmarks against which capabilities are selected is appreciated.

Current measures of fitness: Error, risk and opportunity

The traditional framework for comparing capability development processes is in terms of error, risk and opportunity. The process should allocate resources to projects proportional to the ratio of return to risk. The bulk of the decision-making process in capability development is geared around minimising the error of estimating the return and the risk for individual projects. For example, the Kinnaird review [182] recently resulted in changes to the capability development process to provide more structured roles for DMO and DSTO in assessing and signing off on the schedule and technical risks respectively for every project. This makes good sense from the perspective of error, risk and opportunity.

The current process measures the fitness or value of a project by its promised contribution to predicted future requirements articulated in high-level strategic guidance. This is a purely top-down approach to evaluating capability, since it is measured by the perceived value according to those at the top of the organisational hierarchy. Under a centralised planning framework this makes sense because central decision-makers have a bigger picture view of the organisation, its purpose and its likely future. The greater the potential of a project to deliver decisive capability in future operations, the more resources the process is willing to invest in that project. Resources may be effectively fully committed to an acquisition in this process before the equipment has ever been built or tested, due to sunk costs and an inflexible process acting as a barrier to change. The concept of efficiency in this context refers to the ability to deliver a project to requirements on time and on budget.

To mitigate the risk of a strategy that could lead to lock-in, a key driver in the current process is maximising accountability. The existence of an audit trail from identified future requirement through to the delivered capability is necessary to ensure the risk of misspending taxpayers' dollars is minimised. The audit process tends to be self-reinforcing, since if a project delivers a capability that does not meet the original requirements, one cause that is often blamed is some form of leeway in the current accountability process. The obvious remedy is to subject future projects to greater control. Of course, even within this perspective it is obvious that there is a point where regulating more details of the process will lead to inefficient use of taxpayers' dollars due to the overheads of complying with the process. However, the cost of compliance can be justified as insurance against large deviations from the expected capability a project will deliver, and large deviations have the most detrimental impact for the Minister of Defence, in the media and on morale.

Interoperability in current processes takes a traditional systems engineering perspective on integrating a system of systems. In order to achieve interoperability it is necessary to develop systems in relative independence and then spend additional time and money to integrate the separate systems. In order to minimise the integration costs, the relations between systems are mapped out using tools such as Use Cases to identify functional flows. The precise relations that are required between systems are then specified in detail and this minimal set is implemented. The informal term for the implementation of the integration function is glueware. As the name suggests, glueware holds the system components together, but it also shares another property with its namesake: glueware tends to set [236]. This means there is only one way (or a very small number of ways) in which the system of systems can be organised to perform its minimal set of functions. It also means that as new systems are added to a system of systems, they must find a way to interact with the current agglomeration rather than reorganising the interactions of several component systems. Over time, the integrated system of systems becomes a spaghetti of fixed interactions and to improve the system of systems, engineers are forced to redesign it essentially from scratch.

A new measure of fitness: End user value

There exist alternative frameworks for assessing the fitness of a capability development process. Given the non-stationary threat environment described in Section 8.2.1, the fitness of a project cannot be determined in advance, but relates to its utility to the end user during real operations. This is an important distinction, since it redefines our understanding of efficiency. A project could deliver to requirements ahead of time and under budget, but still have a low efficiency if it has no impact on capability when it is deployed. Similarly, a project that does not meet the original requirements could be highly efficient if it is used during operations by end users in ways the capability developers never envisaged.

An example is the use of the 88mm Flak guns (see Figure 8.3), initially designed solely for anti-aircraft defence, in many other roles including anti-tank, vital asset protection, bunker busting, offensive support, sea targeting and beach landing defence. Rommell most famously used the 88mm Flak gun during the Battle of

Sollum in June 1941 at Halfaya, where they destroyed 123 out of 238 attacking British tanks from concealed dug-in positions. The versatility and effectiveness of the 88mm Flak gun can be explained by the employment of building blocks and coevolution. Having interchangeable components allowed the development of multiple variants that recombined the basic building blocks to produce different capabilities for different roles. The variants then co-evolved in response to the way in which commanders employed them in the field. For example, the early use of the 88mm Flak guns during the Spanish civil war to attack bunkers and troops led to the development of the Sonderanhaenger 201 (special trailer) enabling firing at a low angle without lowering the mounting to the ground. From 1940 the Sonderanhaenger were fitted with shields for protection during direct fire operation. The 1941 Flak variant rotated the mount by 90° to reduce the silhouette. None of these features were required for the original anti-aircraft function, but once end users discovered the utility of the gun in different roles, feedback into the acquisition process enabled the adaptation of the equipment to the new context.

NOTE: This figure is included on page 159 of the print copy of the thesis held in the University of Adelaide Library.

Figure 8.3: 88mm Flak Anti-Aircraft gun in use as an Anti-Tank weapon, after [212].

Once it is acknowledged that the context is non-stationary and unable to be predicted, the utility of centralised futures planning must be questioned. Although the top of the hierarchy has access to a broader spectrum of information, the information has been filtered to reduce complexity and the decision-makers are removed from direct operational experience. Any change in the context will almost certainly be first sensed at “the edge" of the system, in the lowest ranks of the hierarchy. Since this is also where the value of capability is realised, the aggregation of requirements from the edge of the network could provide much more efficient capabilities. Moving power to the edge is seen to be a central issue for force transformation [5], but it is equally applicable to the capability development process itself.

There are currently four arenas where value is determined in a decentralised way that balances competitive and cooperative forces: science, democracy, the justice system and regulated free market economies [61]. For example, democracies and free market economies both rely on signals of preference from end users (voters and consumers) to reinforce strategies that provide value to the whole system. Development in these arenas occurs through evolution rather than central control. A necessary condition for "edge organisations" is the ability to collect and aggregate signals of preference from end users. There are many mechanisms that could facilitate this process within capability development, ranging from surveys and referendums through to a level of direct purchasing power.

From this perspective, efficiency is no longer a primary objective. The focus on optimisation within a traditional systems engineering context can result in optimality under a very specific set of conditions, but systems that are brittle to changes in the environment or underlying assumptions. The optimal solution on a fitness landscape is all too often on the edge of a cliff. For example, the optimal racing line for a car around a barrier is immediately adjacent to the racing line that hits the barrier. In this case, the optimum exists on a discontinuity that leaves no margin for error.

Optimisation, combined with our first definition of efficiency, results in methodologies based on standardisation and eliminating redundancy. An example of this thinking in the DCDM is the justification of a centralised filtering process "to reduce the amount of unnecessary work" [22, Section 2.23]. This approach is ideal for industrial manufacturing, where production lines transform tightly controlled, constant inputs into constant outputs. However, a side effect of increasing the efficiency of a system in this way is that the elements become tightly coupled, which in complex systems dramatically increases the risk of catastrophic failure [246]. These failures are caused by the amplifying effect of tight coupling, which allows small perturbations to cascade through the system.

Insights from complex systems science propose exactly the opposite of the traditional standardisation and efficiency framework: diversity and apparent redundancy are essential elements of a robust and adaptive system [33, 139]. From this perspective, it is natural to develop a portfolio of possible capabilities rather than pursuing a single option. Different options could be funded in parallel at a low level, with end users receiving small numbers of rapid prototypes. Value is assessed in the most realistic environment available by users, and competing capabilities have their funding grow at the expense of others based on relative assessments of fitness. Successful capabilities survive, and are reinforced by larger scale development and deployment across the ADF. Weaker programs must die out when they can no longer successfully compete for their funding. Because the capability development process would deliberately maintain a level of diversity, as the context changes weaker programs (as assessed in the old context) may become more relevant, and as their funding grows with the corresponding increase in fitness, the capability could be scaled up rapidly, when compared to a capability development process that commits to selecting a single capability option optimised for the old context. Evolutionary approaches that incorporate strong real-world feedback deliver capabilities that are characterised by their effectiveness, rather than efficiency.

Integration would also acquire a new emphasis. Rather than developing glueware, integration would set common languages and interface standards so that systems could self-organise into many different working configurations. Here, self-organisation refers broadly to an increase in organisation that does not follow an explicit instruction template and is not centrally directed.

The Internet is a paradigmatic example of how the specification of minimal interface standards delivers enormous flexibility in the functionality and topology of interactions. As new components are added to a system, the topology of interactions can reorganise into a new stable pattern. As legacy components are replaced, the interactions with other components can gradually shift from the old to the new. This has the added benefit of providing redundancy and robustness. Redundancy in turn allows new capabilities to be more innovative and less risk averse, since the existing components provide safety backup until the new components are proven in real-world contexts.

8.3.2 How complex is capability development?

All systems contain a fundamental tradeoff between complexity and scale [32]. Here, complexity is measured by the distinguishable variety in the system, while scale means to multiply or repeat a component of the system. Intuitively, this tradeoff states that generating large scale effects is only possible by sacrificing the variety of the components of the system. In this section, the current emphasis on large scale acquisition is contrasted with an alternative complex acquisition approach.

Current approach: Large scale acquisition

The current approach to capability development is a large scale approach. Many projects are a significant percentage of the defence budget, and involve the mass production of standardised equipment which is then deployed across large parts of the ADF. For example, the Steyer rifle, the ASLAV armoured vehicle and the RAVEN communications sub-system have been produced in several standard variants at large scale. Again, the motivation for large scale acquisition is efficiency. The economic growth of the industrial age was a result of the efficiencies of large scale transformation of resources into goods and services.

The large scale nature of most acquisitions demands a rigourous capability development process, since errors are amplified by scaling, and can significantly threaten the capability of the ADF. The requirements for rigour and long term planning contribute significantly time lags and inflexibility in the capability development process.

A New approach: Complex acquisition

Section 8.2.1 identified increased complexity as a challenge for the ADF. The draft Australian Army Complex Warfighting concept [181] outlines a new type of organisation to cope with increased complexity. Smaller, austere, semi-autonomous teams with modular organisation are envisaged. These teams use swarming tactics and devolved situation awareness to operate as self-reliant teams that aggregate to achieve larger scale effects through local coordination rather than central control.

This kind of force structure places very different demands on the capability development process. Whereas a conventional force demands standardisation so that effects can be synchronised and the output of units is predictable, complex warfighting teams will be heterogeneous and have varied demands for materiel. An industrial age approach to capability development tailored to large scale production of standardised materiel is unsuited to meeting the fine scale but complex demands of the new force structure.

Complex acquisition requires the delivery of tailored solutions to different users, which increases variety. It also implies closer feedback between end users and the acquisition process. There is still room for large scale production, but this would be driven by end user demand rather than strategic planning. An alternative to conventional large scale engineering – complex systems engineering – was described in Section 7.3.

8.3.3 How are capability gaps identified?

The identification of capability gaps plays a crucial role in establishing the business case for large investments in materiel, providing an audit trail between projects and unmet military strategic needs. In this section, we consider the implications of approaching this problem in reverse, starting with the force in being rather than the high level strategic need.

Current philosophy: Inducing requirements

While comparing two capabilities against one possible future is difficult, comparing and deciding between alternatives against an open-ended ensemble of possible futures is a truly “wicked” problem. Rather than tackle this problem directly, the current process addresses the simplified problem of a number of generic point scenarios that include representative elements of future operations. This idea is illustrated by the first recommendation of the recent Kinnaird Review [182] of procurement in the ADF:

> *Recommendation 1*
>
> Defence should present to government the following information in a succinct form on an annual basis:
>
> - an assessment of the types of contingencies Australia might face in carrying out the strategic tasks endorsed by government in Defence White Papers;
> - advice on the military force required in each contingency and the capacity of the ADF to apply this force now and in the future; and

- advice on capability to be sustained, acquired or retired to ensure this can be achieved at acceptable cost.

As described in Section 8.2, these point scenarios are identified broadly in the high-level strategic guidance. The most common analytical basis for assessing capability gaps are the Australian Illustrative Planning Scenarios (AIPS). The AIPS are designed to cover the spectrum of operations in dimensions such as intensity of conflict, type of adversary, geography and environment, and operational objectives.

For each context, the force structure and capabilities required are induced. The capabilities are compared to the Force In Being (FIB) and current capability projects to identify capability gaps. Each context reveals a different set of capability gaps. When aggregated, the capability gaps may produce an inconsistent or conflicting set of requirements. For example, mobility in high-intensity conflict may require a high degree of survivability, while mobility in peacekeeping operations may require a vehicle that does not intimidate or discourage interaction with the local population. The aggregated requirements list is highly dependent on the details of the scenarios chosen, and any bias towards one type of scenario will be reflected in the requirements list. This also introduces the need to rank the relative importance of the point scenarios in order to prioritise the capability gaps, which is currently done centrally and at the highest level.

A new philosophy: Deductive exploration

In contrast, a deductive approach starts with the FIB. Human creativity is harnessed to invent new force structures and capabilities that can be added to the FIB, within some resource limit. A new force structure can then be tested against the AIPS, and is only considered if it may outperform the current force structures in all scenarios. Unlike induction, the deductive approach is logically valid. More importantly, the deductive process begins with a coherent force structure that includes the FIB, so the requirements do not contain inconsistencies. If capabilities are viewed as building blocks that can be reorganised to fulfil different missions, the number of useful ways they can be organised provides a measure for the potential number of different operations a given set of building blocks could perform. In other words, this measures the force's flexibility, or equivalently its variety. This enables a comparison of the reach of different force structures within an open-ended ensemble of possible futures, which is not possible within an inductive framework.

Because the deductive process starts with existing components that may be reorganised, it is likely to result in better reuse. Interoperability and integration now naturally fit with the new focus on self-organisation, rather than generating glueware. Recombining the interactions between existing components with small modifications, rather than acquiring new systems from scratch, can satisfy new requirements quicker. This is reminiscent of the way nature opportunistically reuses building blocks for new functions. For example, the evolution of limbs from fins shows how the lobe-finned relatives of early tetrapods produced weight-bearing fins during the Devonian period [84]. This is thought to have eventually changed the function of their paired pectoral and pelvic fins from providing underwater stabilisation and lift to moving across land.

8.3.4 How are capability development projects controlled?

Approaches to controlling the implementation of capability development decisions can be broadly characterised as either centralised or decentralised. Neither approach is intrinsically superior, as each approach is useful for different contexts. The majority of structured human activity is organised similar to a centrally controlled hierarchy. Interestingly, most other living systems are better described by distributed self-organising networks. In this section, the current centralised process is contrasted with an alternative decentralised model for capability development.

Current approach: Top-down centralised control

The current approach to capability development is a predominantly top-down directed process. All major projects must demonstrate a link to centralised strategic guidance. This provides a rational audit trail for decision-making, which maximises accountability. Projects are then allocated with the cost, schedule and content of the project pre-defined. The posting cycle of 2-4 years means that program managers will generally not control a project from start to finish, since the average project is significantly over 5 years. As caretakers, the incentive structure for program managers is geared towards minimising the risks of not meeting the cost, schedule and content goals. Program managers have a limited ability to change these parameters to avoid unanticipated threats or exploit new opportunities.

Centralised control linearly transforms requirements into implementations, since it requires sequential, synchronous decision-making. Centralised control is usually implemented as a hierarchical control structure. It is most effective for problems with bounded complexity, when prediction and precise control are both possible and desirable. The advantages of hierarchies (see for example [329]) include the ability to exercise tight constraints on the behaviour of large numbers of people from the top down. The ability of different levels of the hierarchy to solve problems of different scales allows the system to apply course-graining filters and operate at multiple levels of complexity, provided cross-level interactions are minimal. The major limitation of hierarchies is they cannot cope with significantly greater complexity than their central node, which is usually an individual. This means there is a limit to complexity, beyond which the hierarchy is no longer the most effective way of organising human activity [25].

An insightful exposition of centrally controlled hierarchical organisation is given in Bar-Yam's latest book [34]. The example is lengthy but highly relevant to understanding the limitations of centralised control.

> In the Soviet Union tremendous effort was devoted to planning the economy. There was a general five-year plan, and then there were detailed one-year plans that were broken up further into one-month plans. They used a form of computerized scientific management, as well as a careful negotiation process between individuals who were responsible for individual enterprises in the system. In the one-year plans, the flow of materials, products, labor and money was directly specified for each product within each enterprise. Not only was what went in and

out specified but also where it came from and where it went to. On a daily basis (and then weekly, monthly and yearly) the flows of money were monitored by the banking system so that they corresponded to the plans. The prices were set centrally so that the flows of money corresponded to the flow of materials, products and labor. The planners were well aware of the U.S. free market system and they viewed it as wasteful. Planning, they believed, would lead to increased efficiency due to an elimination of wasteful duplication of effort. In the free market system there are multiple companies doing the same thing. This repetition of effort seems to planners to be a waste of labor and capital.

How well did the planned system work?

In a supermarket in Moscow, the total number of possible foods you might find was only roughly a hundred... Most of the time even these foods were not available. It was a system where scarcity was the rule. People had to be satisfied with what there was, not what they wanted... Waste was very high, 20-50%, even though the items were very scarce. A substantial fraction of fresh fruit rotted in warehouses... Waiting in line and shopping generally took a substantial fraction of people's time and a significant fraction of income was spent on food.

This was the main food system of the Soviet Union... The farmers' markets were the main source of additional food options, though at significantly higher prices.

Contrast this with the U.S. food supply system at the time. American supermarkets in this period were stocked with well over 10,000 products (today nearly 40,000), selected from over a hundred thousand possible products by supermarket owners... Food of various types, prices and qualities was available at essentially all times (24/7) and in all locations. The economy as a whole was and is consumer-limited rather than supply-limited, so that advertising is necessary for sellers to promote their products.

There is a direct connection between the failure of the Soviet food system to provide adequate improvement and the collapse of the Soviet Union. The person "in charge" of agriculture in the USSR from 1978-1985 was Mikhail Gorbachev, before he rose to become General Secretary in 1985. His college degree was as an agronomist-economist. The ineffectiveness of the agricultural system led to Gorbachev's efforts to change the Soviet system and might be considered among the immediate causes of the collapse of the USSR. The leaders of the USSR were very aware of the comparison of their effectiveness as measured in comparison with the U.S. and other countries. Thus, we would be well justified in saying that the inability to perform the complex task of food production and supply, as compared with the effectiveness in other places, contributed to the downfall of the centrally planned economy of the USSR.

The collapse of the Soviet Union, the free markets in China, the change

> of many governments from dictatorships to more democratic systems and the implementation of [Total Quality Management] in corporations all point to the inability of central control to effectively manage the complexity of modern social organizations in the face of complex external forces and demands.

The planning process for the Soviet economy has a similar perspective to the current capability development process. The underlying justification is that analysis and rigorous planning results in more efficient utilisation of resources and better decisions. The problems experienced within the Soviet economy also have analogs within capability development: end users have little choice and little say, are forced to wait for costly equipment (the media continually reports budget and schedule over-runs), ammunition and consumables are scarce, and the system does not appear responsive to changing needs. The central planners have observed similar symptoms of inefficiency from "capability decisions that time has rendered irrelevant, ineffective or sub-optimal".

A new approach: Self-organising capability development

By viewing program managers as entrepreneurs, an alternative approach to capability development can be advanced. An entrepreneur has control over the niche their business occupies. They define the need their project will fulfil and the solutions they provide. Rather than being rewarded on the promise of their project, an entrepreneur is rewarded for delivering successfully to their clients. An immediate implication of this model is that some project deliverables must occur at a faster rate than the posting cycle, so the entrepreneurs can receive feedback.

If the capability development process is viewed through the lens of business analysis, an alternative to top-down definition of capability needs becomes apparent. The requirements we described in Section 8.2.1 for capability development are analogous to the requirements for the set of all suppliers in an economy. Suppliers want to minimise the cost of producing goods and services while maintaining agility with respect to shifts in consumer tastes and demands. While short-term trends in demand can be identified, long-term changes in consumption are unpredictable. Central planning to manage the evolution of supply to meet demand has already been shown to be unsuccessful in the case of the Soviet food production and distribution system [34]. Centralised control is unable to scale up to the complexity of this kind of problem. In a non-stationary and disruptive dynamic context one type of agent drives the self-organisation of the economy to meet demand: the entrepreneur.

An entrepreneur is rewarded for identifying and fulfilling a consumer's need for less than or equal to the price they are willing to pay for it. The entrepreneur has control over the niche they fulfil and interacts with other suppliers in both competitive and cooperative relations, depending on the degree to which their neighbouring suppliers' goods and services substitute or compliment their own product. Whether or not individual entrepreneurs adapt to changing tastes, the economy as a whole is still adaptive.

The entrepreneurs whose goods and services are most demanded receive the most revenue, and the proportion of revenue that is reinvested reinforces the growth of the entrepreneur's company. Entrepreneurs that provide no value to consumers soon go out of business. The large personal rewards for success provide the incentive for entrepreneurs to take risks and identify latent regions of value.

The larger and more valuable an unsatisfied need becomes, the greater the rewards that are available to a new entrepreneur. In this way, the economy effectively "auctions off" new niches to the first entrepreneur who can identify and supply goods and services for a given cost. The value of new goods and services is always assessed by consumers.

This entrepreneurial model for capability is fundamentally decentralised. Program managers are given greater autonomy rather than being centrally controlled. Decentralised control allows parallel, asynchronous decision-making, which is non-linear and can scale effectively with increasing complexity.

We are not suggesting that program managers should be subjected to all the extremes of market forces. However, a degree of market pressure could change the incentive structure, such that the capability development process is more adaptive. There are many mechanisms for the attribution of credit to entrepreneurial program managers, and only a subset of these are financial.

Another important caveat is that free markets operate best (meaning that certain undesirable characteristics, such as not accounting for externalities or over-utilising common resources, are reduced) with a degree of central regulation. Constraints are required to ensure negative externalities are paid for within the system and that self-interest does not dominate the interests of society. In the context of the capability development process, market regulation corresponds to constraints on the freedom of action of entrepreneurial program managers. This activity in [33] is described as the creation of a planned environment that fosters learning by doing and enables unanticipated advances using evolutionary engineering.

In capability development, an entrepreneurial model for managing capabilities could provide the engine for adaptation and agility in capability. In order for the model to succeed, the capability development process must be modified to operate as an effective marketplace.

8.4 Summary

CDG has identified agility and adaptability as important characteristics of the capability development process. We have shown why this is important by linking changes in the threat environment to the need for a flexible capability development process. A detailed examination of the symptoms, issues, complexity and metrics for improvement provided by the client established some of the problems with the current process and identified the requirements for a new capability development process.

In Section 2, four new perspectives were contrasted with the current capability development process. By transferring some decision-making power to the edge of the organisation, capability development can be more agile to changes in context.

By incorporating complex acquisition methods, the process can generate a greater variety of capabilities, which increases the force's ability to cope with complexity and evolve novel solutions. Through a deductive exploration of possible force structures, projects can be selected in the face of great uncertainty. By changing the incentive structure and increasing the autonomy of program managers, the capability development process can adapt to avoid threats and exploit opportunities.

The new perspectives we have surfaced in this chapter provide insights about some limitations of the current process, and have suggested new directions that may overcome systemic issues in the capability development process. A detailed analysis of those directions is beyond the scope of this chapter, but would appear to be an obvious direction for further research. In particular, the issue of "market design" requires a synthesis of deep understanding of the functioning of successful markets, and the specifics of the capability development environment, before specific recommendations could be made.

We are not advocating that the old metrics, production and control structures be discarded. Instead we would like to emphasise that the traditional perspectives are incomplete. We should aspire towards a dynamic balance between the current process and new mechanisms that increase adaptability. If agility and flexibility are truly desired, traditional approaches must be augmented with greater end-user participation, complex acquisition methods and greater autonomy for program managers.

Chapter 9

Conclusion

This final chapter summarises the results of my thesis. More importantly, I reflect on the relationships between chapters, showing how these connections allow us to transcend the false dichotomy between inquiry and application, and the arbitrary boundaries between disciplines. The conclusion of my thesis is that the best way to understand, intervene in and design complex systems is with a multidisciplinary approach.

9.1 Ontogeny of This Thesis

In Greek, *θέσις* literally means position. To establish a position, it is necessary to arrange one's research into a logical progression that tells a story whose assumptions are reasonable and whose conclusions are justified. In this thesis, the story starts with a seed of doubt, found at the core of rotting fruit fallen from the tree of model-based approaches. It begins in earnest by establishing new roots in philosophy, matures by growing a strong trunk of theory, and culminates in the fruits of my labour, two ripe and juicy real-world applications.

This tale lies in stark contrast to the actual sequence of events from which it was constructed.

My candidature commenced with a set of seven applied complex systems projects that crawled along roughly in parallel. These applications provided me with a broad and healthy dose of experience, but their only tangible contribution to this thesis was the application in Chapter 1 – an example that I believe demonstrates the misapplication of complex systems. The rest of the thesis unfolded by chapter almost exactly in reverse order, as my interests migrated towards more theoretical concerns, and then into philosophy. It was only towards the end of the write-up that the unifying theme of multidisciplinary conceptual analysis became apparent. But even this account belies the non-linear nature of research, since formalising the theory induced changes to the applications, while developing the philosophy unleashed still greater modifications on the later chapters. Although inconvenient, the frequent revisions signified firstly that my inquiry was generating meaningful distinctions, and secondly that the parts were strongly related.

That philosophy is necessary for sensible applications.

That theory needs to be grounded by worked examples, and informed by more general principles.

This sounds obvious, but then why is breadth systematically deferred in favour of depth in academia? Do the limits to cognitive capacity necessitate organisation into intellectual stovepipes? It gradually dawned on me that not only could my thesis refute such a conclusion by counter-example, but that this was the deepest theme of my research. This chapter concludes by re-capping the contributions of each chapter. I then focus on the multi-disciplinary relationships between the chapters. They demonstrate that there are insights that cannot be found through specialisation – emergent properties that do not fit within the scope of disciplinary stovepipes. Finally, I make some observations on future research in multi-disciplinary approaches to complex systems design.

9.2 Summary of Results

Recall the questions in Section 1.4 that motivated this thesis.

1. How do models represent the underlying system of interest?

2. What is a systems approach?
3. What sort of insights are unique to a complex systems approach?
4. Are there alternatives to model-based applications of a complex systems approach?
5. How are the concepts of complex systems interrelated?
6. What is emergence?

These questions are used, in the following subsections, to review the contributions of my thesis.

9.2.1 How do models represent?

The first question was answered in Chapter 2 by developing a general definition of representation:
Definition 9.1 (Representation). *A triadic relation between a model, entity and agent. The model substitutes for at least one entity, shaping the behaviour of at least one agent.*

This identifies the crucial role of the agent in constructing representations, and how models must affect the agent's predisposition to behave. However, many more detailed questions on representation were also addressed. There are three types of pure ground for representations: iconic, indexical, and symbolic. The three types of external models are formal systems, analogs and language. The general nature of my definition of representation was demonstrated by considering representation in metaphysics, military theory and systems theory.

In contrast to external models, internal models give rise to an infinite regress, unless the interpretant is ultimately a change in external behaviour. Internal representation underwrites explanations of the purpose of the brain in classical and connectionist cognitive science, but this assumption has been challenged by a growing number of anti-representationalist arguments. Analysis of this debate in the context of Peirce's triadic relation revealed the insight that rather than the dominant metaphor of representation-as-mirroring, a more accurate account is of representation-as-inference. Finally, Chapter 2 used the theory of representation to distinguish between agents and objects. In this account, agents and representation are inseparable: agents necessarily represent their environment, and representation always involves an agent.

9.2.2 What is a systems approach?

The second question was also answered with a definition:
Definition 9.2 (System). *A system is a representation of an entity as a complex whole open to feedback from its environment.*

A history of the systems movement provided the context in which the different elements of this definition could be appreciated. The definition incorporates ten essential features that are characteristic of the systems approach. The systems movement was cast as a reaction to the excesses of the Cartesian analytic method, a reductionist form of mechanism that was inadequate for problems of complexity, social science and management. However, systems science is also different from holism, because systems are explained in terms of their components, whereas holism appeals to intuition, which leaves components unanalysed. Another way of answering the question "what is a systems approach?" was provided by viewing the systems approach as the critical revision of modelling assumptions. Eight common modelling assumptions were listed, assumptions that systems research has challenged by developing new techniques for capturing and analysing organisation.

General systems theory, cybernetics, systems analysis, systems engineering, soft systems and complex systems approaches were described and distinguished. The most significant difference between the many systems approaches is the difference between hard and soft systems approaches, because it is a metaphysical choice between belief in the one and the many.

9.2.3 How is complex systems unique?

Complex systems studies the mechanisms of dynamic pattern formation over multiple scales. Emergence, self-organisation and adaptation are the main themes of complex systems. Of course, these themes are not unique to complex systems – they are important in many of the systems approaches. The question of how complex systems differs from alternative systems approaches was not explicitly answered, but it is easy to infer from the descriptions of the different systems approaches in Chapter 3.

Perhaps the most notable feature is that, of all the systems approaches, complex systems is the most tightly integrated with the natural sciences, due to an emphasis on explaining mechanisms in natural systems. In contrast, systems engineering usually has more pragmatic concerns. Meanwhile, systems analysis and is more analytic, and consequently focuses on extracting from a system of interest just the information that is relevant to decision-making. This can often be achieved from phenomenology ie. without a detailed understanding of mechanisms. In relation to science, systems analysis adheres most closely with management science, due to their mutual interest in decision-making. Cybernetics was explicitly behaviourist and independent of science for its foundations. Soft systems and critical systems approaches draw more from sociology than from the natural sciences. And while the research agenda of complex systems is largely contained in the enthusiastic visions of the early general systems theorists, complex systems has been less abstract and less concerned with unified theories and general principles.

Generally, the systems approaches place a strong emphasis on the information sciences (information theory, control theory and computer science). However, the agent based modelling technique is distinctive of complex systems. ABMs are uniquely able to represent the emergence of collective behaviour in a discrete time simulation, where decision-making is distributed amongst self-organising agents following simple rules.

9.2.4 Are there alternatives to model-based applications?

The alternative that was advanced in Chapter 4 is conceptual analysis. Conceptual analysis is a multidisciplinary approach that utilises the exact results of complex systems without building quantified models of the system of interest. It was shown that mathematically precise theory in conjunction with the critical examination of assumptions provides a powerful method for arriving at insights using low resolution models, without the need for detailed quantitative input. Three issues for conducting multidisciplinary research, namely standards, inconsistency and justification, were addressed. It was found that conceptual analysis needs to examine assumptions and inconsistencies in multidisciplinary models, and make small scale changes to systems in context, using evolutionary mechanisms to improve system performance over time. Conceptual analysis shifts the focus of modelling from producing rigourous analytically justified designs, towards generating novel ideas that can then be implemented and assessed in context.

Part III exhibited two applications of conceptual analysis. The first example analysed asymmetric warfare, where the insurgency's tactics combine to increase fine scale complexity. The law of multiscale variety explains why this strategy is effective against conventional forces. Recent military theory was supported by and extended using complex systems concepts, such as adaptation and complex systems engineering. The second example engaged the Australian Capability Development Group in thinking about how to make the capability development process more adaptive. Using a structured framework of questions and doctrinal information on the current process, a number of assumptions were critically examined. This gave rise to a new measure of fitness based on end user value; a new approach to acquisition to facilitate variety and tailored capabilities; an alternative philosophy for justifying expenditure; and an entrepreneurial model for program managers. These applications built low resolution conceptual models, which were used to generate new ideas in force design and suggestions for change in the current capability development process.

9.2.5 What are the conceptual interrelationships?

Chapter 5 answered this question in several different ways. Firstly, an illustrative story was told about a cyclical process that generated increasing complexity through self-organisation and emergence. This is the sort of story that can be found in many popular books on complexity. The next answer was given as a series of definitions formalised using information theory. These definitions were supplemented with a number of examples, which illustrated that each concept was a useful distinction in its own right. Finally, this more precise understanding of the language of complex systems was reassembled in diagrammatic form to provide a representation of the important interrelationships between these concepts.

Chapter 5 also made several deep connections. The formula

$$\begin{aligned}\text{mutual information} &= \text{receiver's diversity}\\ &- \text{equivocation of receiver about source}\end{aligned}$$

together with the identity of predictive information with excess entropy leads to the relation

$$\text{predictive information} = \text{diversity} - \text{non-assortativeness}.$$

The predictive information expresses a recurring theme within complex systems. In the literature, this has been less formally explained as a tradeoff between exploitation and exploration, or between being adapted and adaptable. Connections were made to information transfer in graph theory, robustness in self-organising systems, increases in physical complexity through evolutionary processes, and self-referential processes such as shortest path formation by ants. Each process was shown to be a balance between diversity and equivocation.

9.2.6 What is emergence?

The answer to this question, given in Chapter 6, was once again able to be formulated as a definition:

Definition 9.3 (Emergence)**.** *Emergence is the process whereby the assembly, breakdown or restructuring of a system results in one or more novel emergent properties.*

However, the definition of an emergent property proved to me more interesting, once the framework of scope, resolution and macrostate/microstate was formalised.

Definition 9.4 (Emergent property)**.** *A property is emergent iff it is present in a macrostate and it is not present in the microstate.*

It was shown this definition could be narrowed further to novel emergent properties, since changes in resolution only generates weak (apparent) emergent properties.

Definition 9.5 (Novel Emergent Property)**.** *A property is a novel emergent property iff it is present in a macrostate but it is not present in any microstate, where the microstates differ from the macrostate only in scope.*

This definition is more specific, because it only counts emergent properties when they are not a property of *any* microstate with smaller scope. This definition provides the key insight of Chapter 6 – that emergent properties are coupled to scope, not level. Defining emergent properties precisely allowed alternative interpretations. Equivalently, an emergent property is: non-local; spatially or temporally extended; or a difference in local and global structure.

It was shown that naissance emergence, the source of novelty, cannot be predicted with certainty until it has already occurred. This is a fundamental limit for formal systems with respect to emergence. The definition of an emergent property was used to describe an alternative process for defining the boundary of a system. It was argued this process was less subjective, depending only on the choice of properties belonging to the system.

9.3 Interconnections

Many of the topics of this thesis were reflexive. Contributions to the theory of multidisciplinarity were performed in a multidisciplinary way. The theory of representation provided an understanding of the representations I employed. As I developed a theory of adaptation, I was also very conscious of how my theory was adapting. My theory of emergence literally emerged – it was not until my chapter had a certain scope and structure that I was able to recognise its unique contribution. This is more than a superficial play on words; it identifies important relationships between concepts with themselves. Often, this allowed the contributions to be interpreted on multiple levels. For instance, representation was necessary for modelling the behaviour of agents that construct representations, but also applied to the modelling process itself.

There were several strong links made between representation and systems. Systems theory contained several candidate theories of representation; while systems were defined as representations in Definition 3.4. The systems concept of 'agent' was also intimately involved in representation, because representation is a triadic relation involving an agent; and agency requires the ability construct and use representations. These connections were surprising, because the two topics have developed almost completely independently in the literature. To my knowledge, the inseparable relation between agency and representation has not been previously published.

Information theory provides an important interpretation of representation, including concepts such as mutual information and predictive information, which can in principle quantify how well a model substitutes for an entity. Representation was necessary to explain adaptation in Section 5.7, because both learning and evolution "encode information" about the environment in models. The connection here is that in order to reinforce useful changes in an adaptive process, environmental feedback must be evaluated with respect to an expectation of the effect of the change – an expectation which is derived from a model. Emergence as predictive efficiency also directly relates to the efficiency of representational models. Additional significant connections between information theory and representation clearly exist, however they are a topic for future research.

The systems movement has always been closely related with information theory, because of the close parallels between the work of Wiener and Shannon, as well as Ashby's clear articulations of the importance of information theory for complexity and control. While this connection is not novel, the comprehensiveness of Chapter 5 contributes towards understanding how information theoretic representations can be applied to modelling complex systems. By providing a formal mathematical framework, information theory also allows the relationships within complex systems concepts to be explored.

Variety was a recurring theme in this thesis. It was formally introduced in Chapter 3; featured in the account of goal-directed behaviour in Chapter Section 2.6; explained the purpose of constructing representations that shape behaviour in Section 3.3.2; was formally connected to information theory by Shannon's tenth coding theorem; was an essential function for an adaptive mechanism in Section 5.7; extended to the law of multiscale variety in Section 7.2; and used as a measure of complexity in Section 8.3.2.

The fact that emergence has an information theoretic formulation demonstrates a relation between Chapters 5 and 6. However, the deeper connection here is a more personal one. When Shalizi visited Australia to explain his information theoretic definition of emergence to the CSIRO interaction task on emergence, it generated a lot of excitement. However, in order to be appropriately scientific, it was decided to form a sub-team to critically examine the definition and identify its limitations. As a member of this sub-team, I was forced to think about precisely when the equation $e = \frac{E}{C_\mu}$ gave counter-intuitive classifications of emergence. The exactness of Shalizi's formulation meant that this was an insightful process, because it focussed my attention on specific cases that required a more satisfactory explanation. Understanding the information theoretic formulation for emergence allowed me to ultimately move beyond information theory and develop the theory and examples of emergence in Chapter 6.

A common theme of systems thinking is that learning can be obstructed by a conventional adherence to disciplinary boundaries [222, vol. I, p. iixx]. This theme was taken a step further in my thesis by developing a number of arguments towards multidisciplinarity. In Chapter 2, the multidisciplinary theme was introduced, and a multidisciplinary analysis of representation was performed. A characterisation of the systems movement as being broader than any conventional discipline was discussed in Section 3.3, continuing the multidisciplinary theme. In Chapter 4, the multidisciplinary approach to complex systems analysis was articulated. This drew on the previous two chapters. The modeller and the model formed a triadic relation with the system of interest. Conceptual analysis was described as a way of applying complex systems theory, and therefore conceptual analysis is a systems approach. The characterisation of the main systems approaches in Chapter 3 provided a context within which the contributions of conceptual analysis could be situated.

Chapter 5 was oriented towards interdisciplinary communication, providing representations of complex systems concepts in mathematics, language and graphical form. Chapter 6 provided support to the case for multidisciplinary research. By emphasising the importance of scope, emergent properties necessitate the development of techniques with sufficient breadth, rather than additional depth. The existence of emergent properties provides legitimation for broader systems approaches that complement specialised scientific disciplines.

The applications in Part III put the multidisciplinary approach into practice on real world problems. Significant novel insights were revealed through broad conceptual analysis. The assumptions of current perspectives were critically examined, and often found to be focussing on a narrow range of solutions. For example, solutions to the challenges of asymmetric warfare are still largely focussed on concentrating large scale effects, when the ability to cope with fine scale complexity is needed. And the capability development process continues to attempt to achieve tighter control on the delivery of projects to requirements on time and on budget. What is really desired though, is not greater efficiency, but greater effectiveness, and the obsession with efficiency actually has harmful consequences for effectiveness. Part of the obsession with efficiency derives from the quantitative approach to modelling, which by quantifying fitness inevitably reduces effectiveness to efficiency. In contrast, conceptual analysis is designed to uncover effective regions of state space, and is a more appropriate use of analysis in the search for novel design and intervention options.

In summary, the components of this thesis are strongly related, in spite of their breadth. The chapters span philosophy, theory and application; they connect mathematics, science and engineering. Their collective contribution is multidisciplinary. Together, they show that the best way to understand, intervene in and design complex systems is with a multidisciplinary approach.

9.4 The End is The Beginning is The End

In my beginning is my end . . . In my end is my beginning.

East Coker by T. S. Eliot

This thesis has covered a significant amount of ground in order to advance the method of multidisciplinary conceptual analysis. Because of its scope, I feel as though I have only scratched the surface of the utility of conceptual analysis. As is always the case with progressive research, this thesis has generated many more questions than it has answered. However, in this concluding section I will limit my attention to six big questions for further research.

1. Can more detailed guidance be provided for conceptual analysis, and is it desirable?
2. For problems where no one person is in charge, how can conceptual analysis be linked with effective intervention?
3. How can multidisciplinarity continue to raise its standards and still be accessible to everyone?
4. How can a precise understanding of emergence be put to work?
5. Is it possible to measure and detect emergence in simulations and in the real world?
6. What is the precise relationship between representation and information theory?

Chapter 4 indicates a new direction for expanding the domain of complex systems application, but much more work remains to be done to flesh out the details. Good examples of conceptual analysis exist, but they need to be systematically compared to identify the general features of conceptual analysis, so that more detailed guidance for future applications can be provided. The second part to this question asks how much guidance is a good thing? Constraints are always necessary because they focus energy so that it can do useful work. But if the conceptual analysis approach is overconstrained, if it becomes a recipe book or algorithm for generating novel designs, then it will in fact not be producing novel designs.

The second question picks up on a limitation of the applications in this thesis. A full application of conceptual analysis would have implemented some changes in the real system, measured their effect in comparison with the expected effect (provided by the conceptual model), and reinforced successful variation to begin the evolution of new designs in context. The applications in [34] also end with conclusions of the analysis stage, rather than with the results of real world interventions. This is in part because the problems that multidisciplinary conceptual analysis is especially suited to are never owned by a single stakeholder. Consequently, effective intervention requires coordinated changes to be made by multiple agencies, and the likelihood of coordination diminishes in direct proportion to the number of agencies involved. This perhaps the most important question for conceptual analysis to address in order to have a positive impact.

The third question is an open question that was stated as a paradox in Chapter 4. There is clearly a need for multidisciplinarity, and this drives the need for coherence on the meaning of terms. Even the meaning of interdisciplinarity and multidisciplinarity cannot be agreed upon. But unlike disciplines, multidisciplinarity cannot force agreement by walling itself off from other discourses until isolation through jargon and other barriers to entry breeds conformity. How can multidisciplinarity develop depth of knowledge and yet remain accessible to all the disciplines it supports?

From a systems perspective, the most important individual contribution of this thesis is the precise formulation of emergence. Emergence is *the* central systems concept, and yet its explanation in terms of levels has been an impediment to a genuine understanding of emergence. Now that an alternative account exists, how can this be put to work? What changes to systems theory and application follow from Chapter 6? The full implications have not yet been explored.

The fifth question is also related to this issue. In more concrete situations, in principle one should be able to measure and test for emergent properties. For example, consider a flock of boids [261] with a control parameter α. Suppose that for $\alpha = 0$, the birds move independently of all other boids, while for $\alpha = 1$, the birds move with total cohesion. If flocking is defined as when the autonomy of the flock subsumes the autonomy of the individual boids, then it should be possible decide whether flocking occurs, provided autonomy can be measured or estimated. As α is increased over $[0, 1]$ starting at 0, at some point the average autonomy of the boids will be exceeded by the autonomy of the flock. At this point, the flock has emerged from the structure of interactions between the boids. Measuring and detecting emergence in a toy simulation model is an important step towards putting the new conception of emergence to work.

The language of information theory provides rich yet precise concepts for information, communication, complexity and variety. It is obvious that models that participate in representation relations can be interpreted as encoding information. Adami shows how evolution can be defined as an increase in mutual information between a population and its environment. A similar connection between representation in general and information theory can no doubt be made. This could provide a precise test between competing models as substitutes for the same entity. It would also be possible to compare the information capacity of representations with different grounds – for example analog versus digital models of the same system of interest. In turn, the triadic relation has interesting implications for information theory. A formal treatment of this connection could be expected to yield important results.

So we have come full circle, back to the beginning of the cycle of inquiry. I welcome you to join me exploring the next six questions in the search for big answers.

Appendix A

Glossary

In the interests of clarity, the technical terms I use are defined here.

A.1 Definitions

Definition A.1 (Abstraction). *A description that does not refer to a unique substantial entity in an arbitrarily well defined region of space-time. Some abstractions have no physical instantiation, while others are lossy compressions of the state of a substantial entity.*

Definition A.2 (Adaptation). *A change in organisation that increases the fit between a system and its environment.*

Definition A.3 (Agent). *An entity that constructs and uses representations to shape its own goal-directed behaviour.*

Definition A.4 (Behaviour). *The response of an entity given a particular a stimulus within a specified context. Broad [62] contrasts 'behaviour' with 'quality', as qualities do not logically require a stimulus or a context to be well defined properties of entities.*

Definition A.5 (Boundary). *Demarcation between a system and its environment. Boundaries are a modelling choice.*

Definition A.6 (Class). *A set whose members share a common property.*

Definition A.7 (Communication). *The transfer of information [287].*

Definition A.8 (Complex System). *A system of interest with behavioural complexity that is not captured by conventional modelling techniques. The source of complex behaviour is the non-trivial organisation of interdependent components with emergent properties.*

Definition A.9 (Complex Systems, Discipline of). *A discipline that studies the mechanisms of dynamic pattern formation over multiple scales. Emergence, self-organisation and adaptation are the main themes of complex systems.*

Definition A.10 (Complexity, Algorithmic). *The algorithmic complexity, or algorithmic information content, of a string is the log of the length of the shortest program for a given Universal Turing Machine (UTM) which can print the string. Algorithmic complexity is maximal if the string is incompressible, meaning there is no program shorter (within a fixed constant) than the one that directly prints each bit of the string. A related measure is descriptive complexity.*

Definition A.11 (Complexity, Behavioural). *The amount of algorithmic information (see algorithmic complexity) necessary to describe a system's response to its environment [25]. A related measure is interactive complexity.*
Definition A.12 (Complexity, Effective). *The length of a highly compressed description of the regularities in an entity's state [133]. A related measure is statistical complexity (see Section 5.3).*
Definition A.13 (Complexity, Multiscale). *The amount of (Shannon) information required to describe a system at each scale or resolution. The total multiscale complexity is always constant for a given system of interest, but a tradeoff between fine and large-scale degrees of freedom means that the same components can be organised to exhibit different "complexity profiles" – different functions of complexity over scale [31].*
Definition A.14 (Complexity, Physical). *The time required by a standard universal Turing machine to generate the entity from an input that is algorithmically incompressible (see algorithmic complexity). Also known as logical depth [42].*
Definition A.15 (Complexity, Structural). *A measure of the spatial organisation of a system. It is a topological metric that ignores composition. Structural complexity is dependent on the spatial and temporal resolution of the observer.*
Definition A.16 (Configuration). *The spatial arrangement of a system of interest.*
Definition A.17 (Context). *A context is a class of environments i.e. an abstract environment.*
Definition A.18 (Control). *The negation of environmental disruptions in order to realise a goal.*
Definition A.19 (Discipline). *An open, organised social discourse with a mutual interest in knowledge development.*
Definition A.20 (Distinction). *A separation between an entity and its background. A* criterion of distinction *indicates what we are talking about and specifies its properties as an entity [217].*
Definition A.21 (Dynamics). *Forces that change the configuration of a system of interest. Dynamics are observed indirectly through behaviour.*
Definition A.22 (Emergence). *Emergence is the process whereby the assembly, breakdown or restructuring of a system results in one or more novel emergent properties.*
Definition A.23 (Emergent property). *A property is emergent iff it is present in a macrostate and it is not present in the microstate.*
Definition A.24 (Emergent Property, Novel). *A property is a novel emergent property iff it is present in a macrostate but it is not present in any microstate, where the microstates differ from the macrostate only in scope.*
Definition A.25 (Emergent Property, Weak). *A property is weakly emergent iff it is present in a macrostate but it is not apparent in the microstate, and this macrostate differs from the microstate only in resolution. A weak emergent property is a limitation of the observer, not a property of the system of interest.*
Definition A.26 (Entity). *Any object that has existence in the physical, conceptual or socio-cultural domains [72]. Reference to an entity implies the operation of a distinction that defines it and makes it possible [217].*
Definition A.27 (Environment). *An environment is a substantial (or concrete) context. The environment of a system of interest is everything outside the system boundary in causal contact with the system – it affects or is affected by the system.*

Definition A.28 (Epistemology)**.** *A theory concerning means by which we may have and express knowledge of the world [81].*
Definition A.29 (Feedback)**.** *A process in which the output signal of a system is measured, and the measurement influences the input to the system in order to regulate future output.*
Definition A.30 (Force)**.** *Fundamental quantity responsible for the physical behaviour of matter. The four fundamental forces are gravity and the three forces relevant to elementary particle physics. These are the electromagnetic, the weak and the strong interactions [299].*
Definition A.31 (Formal System)**.** *The elements of a formal system are [327]:*

- *A finite set of symbols.*
- *A grammar with a decision procedure capable of parsing well formed formulae.*
- *A set of axioms that are well formed formulae.*
- *A set of inference rules.*
- *A set of theorems. This includes all of the axioms, plus all of the well formed formulae that can be derived from previously-derived theorems and the axioms.*

Formal systems can be used as symbolic external models for the purpose of representation.
Definition A.32 (Goal)**.** *An achievable end which can be expressed as a deterministic function of the state.*
Definition A.33 (Hierarchy)**.** *When used to model systems, higher levels control (regulate, interpret, harness) lower levels, whose behaviours are made possible by properties generated at still lower levels. [272]. A hierarchy is a partially ordered set [9].*
Definition A.34 (Information)**.** *The reduction in uncertainty [287]. Information can also be interpreted as the 'improbability' of an event, since it is the negative of entropy ie. the logarithm of the reciprocal of the probability [292].*
Definition A.35 (Knowledge)**.** *Knowledge is the process of forming, maintaining and using representations.*
Definition A.36 (Law)**.** *A constraint on the set of possible future states.*
Definition A.37 (Level)**.** *A layer in a hierarchy. Levels are populated by entities whose properties characterise the level in question [9].*
Definition A.38 (Light Cone)**.** *In Minkowski spacetime under special relativity, a flash of light defines two light cones. The future light cone expands into space at the speed of light, encompassing every point that could be affected by the flash of light. The past light cone includes all spacetime points from which a material particle could reach the point of the flash before it occurred, and therefore affect it. Any two points in spacetime with disjoint light cones cannot be causally connected.*
Definition A.39 (Macrostate)**.** *A state that has a larger scope, coarser resolution, or both, compared to a microstate.*
Definition A.40 (Measurement)**.** *The detection of change, quantified as a rational multiple of a standard quantity.*
Definition A.41 (Microstate)**.** *A state that has a narrower scope, finer resolution, or both, compared to a macrostate.*

Definition A.42 (Minimal Macrostate). *A macrostate* M^* *is minimal with respect to an emergent property, if the emergent property is present in* M^**, and it is not present in any microstate* μ *with the same resolution and narrower scope (ie. in any proper subset of the components of* M^**).*

Definition A.43 (Model). *A representation of an entity by an agent.*

Definition A.44 (Multidisciplinary Conceptual Analysis). *An approach to real world problem situations, which draws on theorems, models and exact results from both disciplinary and interdisciplinary science, in order to construct low resolution, conceptual representations that facilitate interventions and evolutionary change within the problem situation.*

Definition A.45 (Observer). *An observer is an agent that takes measurements or observations of a system of interest to gain information about it. This information can be communicated to others in the form of a description [26].*

Definition A.46 (Observation). *Observation of an entity is the process by which information is acquired about that entity. [72]*

Definition A.47 (Ontology). *A theory of what exists.*

Definition A.48 (Open System). *A system of interest that exchanges energy, matter or information with its environment.*

Definition A.49 (Organisation). *Interdependence between components of a system in both structure and dynamics. In contrast, a system of interest at maximum entropy exhibits independence between components and is disorganised.*

Definition A.50 (Pattern). *A pattern is organisation identified by an observer in a dynamical system (derived from [93, p. 3]).*

Definition A.51 (Problem, Soft). *A problem, usually a real-world problem, which cannot be formulated as a search for an efficient means of achieving a defined end; a problem in which ends, goals, purposes are themselves problematic.*

Definition A.52 (Process). *A sequence of changes in the configuration of a system.*

Definition A.53 (Property). *Properties are functions from objects to characteristics [68] .*

Definition A.54 (Quality). *A property that is logically independent of the context, in contrast to a behaviour.*

Definition A.55 (Representation). *A triadic relation between a model, entity and agent. The model substitutes for at least one entity, shaping the behaviour of at least one agent.*

Definition A.56 (Resolution). *The finest spatial distinction between two alternative system configurations. If a fine (high) and a coarse (low) resolution representation have the same scope, the fine resolution can distinguish a greater number of possibilities, n, and therefore each state contains more (Shannon) information.*

Definition A.57 (Scale). *In physics, scale refers to a standard for measurement that indicates the length below which no further details can be resolved – physical scale is the inverse of resolution. In mathematics, scale is a transformation by multiplication. The physical usage applies to measurement and observation, while the mathematical usage applies to scaling (see Section 6.3).*

Definition A.58 (Scope). *Scope is defined by a spatial boundary. Spatial is used in the broadest sense of the word to include conceptual and formal, as well as physical spaces, provided the system has a physical manifestation.*

Definition A.59 (Self-assembly). *A process that increases the organisation of a* *system by moving towards equilibrium.*

Definition A.60 (Self-organisation)**.** *A process that increases the organisation of a system away from equilibrium, in contrast to self-assembly.*

Definition A.61 (State)**.** *The complete set of variables that describes a system of interest at a moment in time. The state may or may not include the history of the system of interest. Observing changes in attributes results in a transmission of information. According to quantum physics, every substantial system of interest has a unique state, which can be derived from its quantum mechanical state vector or wave function.*

Definition A.62 (Structure)**.** *Interdependence in potential configurations of a system.*

Definition A.63 (Substantial)**.** *Substantial entities are finite, have a definite spatial and temporal existence, and cannot be universal. Synonyms include concrete and physical [68].*

Definition A.64 (Symmetry)**.** *The existence of different viewpoints from which the system of interest appears the same [11].*

Definition A.65 (Symmetry breaking)**.** *When a whole has less symmetry than its component parts in isolation, it exhibits broken symmetry. The symmetry of any steady state configuration of matter must equal the symmetry of the underlying dynamical equations that govern it, unless one or more symmetries have been broken by the interaction of the parts.*

Definition A.66 (System of Interest)**.** *The object of study (see Kline's [185] Definition 2.3).*

Definition A.67 (System)**.** *A system is a representation of an entity as a complex whole open to feedback with its environment.*

Definition A.68 (Theory-ladenness of observation)**.** *A contemporary term for an idea that perhaps was first emphasised in the philosophy of Kant, and that is captured in the following passage from Reichenbach [260]. "Every factual statement, even the simplest one, contains more than an immediate perceptual experience; it is already an interpretation and therefore itself a theory … the most elementary factual statement, therefore contains some measure of theory". However, observation can be at once theory-laden and theory-neutral, in the sense that the data relevant to a theory T can be collected by someone whether or not they believe in T, or, even, whether or not they have knowledge of T [311].*

Definition A.69 (Variety)**.** *Log of the number of possibilities. That is, the number of bits needed to specify the cardinality of an ensemble. Variety is usually measured by (Shannon) information.*

Bibliography

[1] R. Abbott. If a tree casts a shadow is it telling the time? In C. S. Calude, M. J. Dinneen, G. Paun, G. Rozenberg, and S. Stepney, editors, *Unconventional Computation, UC2006*, York, UK, 2006. Springer.

[2] R. L. Ackhoff. *Choice, Communication, and Conflict: A System's Approach to the Study of Human Behaviour*. Trustees of the University of Pennsylvania, Philadelphia, USA, 1967.

[3] C. Adami. What is complexity? *Bioessays*, 24(12):1085–1094, 2002.

[4] D. S. Alberts, J. J. Garstka, R. E. Hayes, and D. A. Signori. *Understanding Information Age Warfare*. CCRP Publication Series, 2001.

[5] D. S. Alberts and R. E. Hayes. *Power to the Edge: Command, Control, in the Information Age*. CCRP Publication Series, 2003.

[6] S. Alexander. *Space, Time and Deity Volume 1*. Macmillan, London, UK, 1920.

[7] S. Alexander. *Space, Time and Deity Volume 2*. Macmillan, London, UK, 1920.

[8] J. A. Allen and S. L. Kitch. Disciplined by Disciplines? The Need for an Interdisciplinary Research Mission in Women's Studies. *Feminist Studies*, 24(2):275–299, 1998.

[9] T. F. Allen. A summary of the principles of hierarchy theory. `http://www.isss.org/hierarchy.htm`, accessed 5/6/2006.

[10] American Society for Cybernetics. Summary: The Macy Conferences. `http://www.asc-cybernetics.org/foundations/history/MacySummary.htm`, accessed 30/11/2006.

[11] P. W. Anderson. More is Different. *Science*, 177(4047):393–396, 1972.

[12] A. Angyal. *Foundations for a Science of Personality*. Harvard University Press, Cambridge, USA, 1941.

[13] Aristotle. *De Anima*. transl. J. A. Smith, eBooks@Adelaide, originally published 350BC, Adelaide, Australia, 2004.

[14] Aristotle. *Metaphysics*. transl. W. D. Ross, eBooks@Adelaide, originally published 350BC, Adelaide, Australia, 2004.

[15] H. R. Arkes and C. Blumer. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35:124–140, 1985.

[16] D. Arnold. Information-theoretic analysis of phase transitions. *Complex Systems*, 10:143–155, 1996.

[17] W. B. Arthur. Inductive behaviour and bounded rationality. *The American Economic Review*, 84:406–411, 1994.

[18] W. R. Ashby. *An Introduction to Cybernetics*. Chapman & Hall, London, UK, 1956.

[19] W. R. Ashby. Requisite variety and its implications for the control of complex systems. *Cybernetica*, 1:83–99, 1958.

[20] W. R. Ashby. Principles of the self-organising system. In H. Von Foerster and G. W. Zopf Jr., editors, *Principles of Self-Organization: Transactions of the University of Ilinois Symposium*, pages 255–278, London, UK, 1962. Pergamon Press.

[21] W. R. Ashby. Some peculiarities of complex systems. *Cybernetic Medicine*, 9:1–7, 1973.

[22] Commonwealth of Australia. Defence Capability Development Manual. Technical report, Capability Development Group, 2005.

[23] P. Bak. *How Nature Works: The Science of Self-Organised Criticality*. Copernicus, New York, 1996.

[24] T. J. A. Baker, M. Botting, M. J. Berryman, A. J. Ryan, A. Grisogono, and D. Abbott. Adaptive Battle Agents: Emergence in Artificial Life Combat Models. In *SPIE: Smart Structures, Devices, and Systems*, 2004.

[25] Y. Bar-Yam. *Dynamics of Complex Systems*. Westview Press, Boulder, Colorado, 1997.

[26] Y. Bar-Yam. Observer. `http://www.necsi.org/guide/concepts/observer.html`, accessed 8/2/2006 2000.

[27] Y. Bar-Yam. Enlightened Evolutionary Engineering/Implementation of Innovation in FORCEnet. Technical report, Office of the Chief of Naval Operations Strategic Studies Group, 2002.

[28] Y. Bar-Yam. Unifying Principles in Complex Systems. In M. C. Roco and W. S. Bainbridge, editors, *Converging Technology (NBIC) for Improving Human Performance*. Kluwer, Dordrecht, The Netherlands, 2003.

[29] Y. Bar-Yam. When Systems Engineering Fails — Toward Complex Systems Engineering. In *IEEE International Conference on Systems, Man & Cybernetics*, Washington DC, USA, 2003.

[30] Y. Bar-Yam. A Mathematical Theory of Strong Emergence Using Multiscale Variety. *Complexity*, 9(6):15–24, 2004.

[31] Y. Bar-Yam. Multiscale Complexity / Entropy. *Advances in Complex Systems*, 7:47–63, 2004.

[32] Y. Bar-Yam. Multiscale Variety in Complex Systems. *Complexity*, 9(4):37–45, 2004.

[33] Y. Bar-Yam. About Engineering Complex Systems: Multiscale Analysis and Evolutionary Engineering. In S. Brueckner, G. Serugendo-Di Marzo, A. Karageorgos, and R. Nagpal, editors, *Engineering Self Organising Systems: Methodologies and Applications*, Lecture Notes in Artificial Intelligence, pages 16–31. Springer, New York, 2005.

[34] Y. Bar-Yam. *Making Things Work: Solving Complex Problems in a Complex World*. NECSI Knowledge Press, Cambridge, 2005.

[35] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[36] G. Bateson. *Steps to an Ecology of Mind*. Ballantine Books, New York, USA, 1972.

[37] G. Bateson, D. D. Jackson, J. Haley, and J. Weakland. Toward a theory of schitzophrenia. *Behavioral Science*, 1:251–264, 1956.

[38] D. F. Batten. *Discovering Artificial Economics*. Westview Press, New York, 2000.

[39] D. F. Batten. Simulating human behaviour: the invisible choreography of self-referential systems. In *The International Environmental Modelling & Software Society (IEMSS) Conference 2004*, Osnabruck, Germany, 2004.

[40] D. F. Batten. Are some human ecosystems self-defeating? *Environmental Modelling and Software*, in press, 2006.

[41] L. V. Beloussov and V. I. Grabovsky. A common biomechanical model for the formation of stationary cell domains and propagating waves in the developing organisms. *Computer Methods in Biomechanics and Biomedical Engineering*, 8(6):381–91, 2005.

[42] C. H. Bennett. Logical Depth and Physical Complexity. In R. Herken, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 227–257. Oxford University Press, New York, USA, 1988.

[43] H. Bergson. *Creative Evolution*. transl. A. Mitchell, Macmillan, London, 1911.

[44] G. Berkeley. *A Treatise concerning the Principles of Human Knowledge*. eBooks@Adelaide, originally published 1710, Adelaide, Australia, 2006.

[45] I. Berman-Frank, P. Lundgren, Y. Chen, H. Küpper, Z. Kolber, B. Bergman, and P. Falkowski. Segregation of Nitrogen Fixation and Oxygenic Photosynthesis in the Marine Cyanobacterium Trichodesmium. *Science*, 294(5546):1534–1537, 2001.

[46] D. Besomi, editor. *The Collected Interwar Papers and Correspondence of Roy Harrod*. Edward Elgar, Electronic version, 2003.

[47] W. Bialek, I. Nemenman, and N. Tishby. Complexity through nonextensivity. *Physica A*, 302:89–99, 2001.

[48] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity, and learning. *Neur. Comp.*, 13:2409–2463, 2001.

[49] M. H. Bickhard. Emergence. In P. B. Andersen, C. Emmeche, N. O. Finnemann, and P. V. Christiansen, editors, *Downward Causation*, pages 322–348. University of Aarhus Press, Aarhus, Denmark, 2000.

[50] B. Blanchard and W. Fabrycky. *Systems Engineering and Analysis*. Prentice Hall, Upper Saddle River, USA, 1998.

[51] N. Boccara. *Modeling Commplex Systems*. Springer-Verlag, New York, 2004.

[52] G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani. Predictability: a way to characterize complexity. *Physics Reports*, 356:367–474, 2002.

[53] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99:7280–7287, 2002.

[54] F. Boschetti and N. Grigg. Mapping the complexity of ecological models. *submitted to Complexity*, 2006.

[55] K. E. Boulding. General systems theory - the skeleton of science. *Management Science*, 2:197–208, 1956.

[56] K. E. Boulding. Peace Research. *International Social Science Journal*, 29(4):604, 1977.

[57] COL J. A. Boyd. A Discourse on Winning and Losing, 1987.

[58] D. Braess, A. Nagurney, and T. Wakolbinger. On a Paradox of Traffic Planning. *Transportation Science*, 39:446–450, 2005.

[59] D. Braha, A. A. Minai, and Y. Bar-Yam, editors. *Complex Engineered Systems: Science Meets Technology*. Understanding Complex Systems. Springer, New York, 2006.

[60] R. Brandom. Truth and Assertability. *Journal of Philosophy*, LXXIII:137, 1976.

[61] D. Brin. Harnessing Conflict and Competitiveness for Society's Benefit. *American Bar Association's Journal on Dispute Resolution*, 15(3):597–618, 2000.

[62] C. Broad. *The Mind and its Place in Nature*. Routlegde & Kegan Paul, London, UK, 1925.

[63] B. Brodie and F. M. Brodie. *From Crossbow to H-Bomb: The evolution of the weapons and tactics of warfare*. Indiana University Press, Bloomington, USA, 1973.

[64] R. A. Brooks. Elephants Don't Play Chess. *Robotics and Autonomous Systems*, 6:3–15, 1990.

[65] R. A. Brooks. Personal communication, 2004.

[66] R. A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. M. Williamson. The Cog Project: Bulding a Humaniod Robot. In C. L. Nehaniv, editor, *Computation for Metaphors, Analogy and Agents*. Springer, Berlin, Germany, 1999.

[67] M. A. Bunge. *Method, Model and Matter.* D. Reidel, Dordrecht, Netherlands, 1973.

[68] M. A. Bunge. *Treatise on basic philosophy.* Reidel, Dordrecht ; Boston, 1974-1983.

[69] M. A. Bunge. General Systems and Holism. *General Systems*, 22:87–90, 1977.

[70] M. A. Bunge. Levels and reduction. *American Journal of Physiology*, 233(3):R75–R82, 1977.

[71] M. A. Bunge. Systemism: the alternative to individualism and holism. *Journal of Socio-Economics*, 29:147–157, 2000.

[72] M. Burke. Robustness, Resilience and Adaptability: Implications for National Security, Safety and Stability (Draft). Technical report, DSTO, 2006.

[73] F. Capra. *The Web of Life: A New Scientific Understanding of Living Systems.* Anchor Books, New York, 1996.

[74] D. Carment. The Struggle for Peace - Rethinking Intervention. *The Future of War*, 23(2), 2001.

[75] J. L. Casti. Chaos, Gödel and Truth. In J. L. Casti and A. Karlqvist, editors, *Beyond Belief: Randomness, Prediction, and Explanation in Science.* CRC Press, 1991.

[76] J. L. Casti. *Complexification: Explaining a Paradoxical World through the Science of Surprise.* HarperCollins, New York, USA, 1994.

[77] J. L. Casti. Would-be worlds: the science and surprise of artificial worlds. *Computers, Environment and Urban Systems*, 23:193–203, 1999.

[78] G. J. Chaitin. Information-theoretic limitations of formal systems. *Journal of the ACM*, 21:403–424, 1974.

[79] G. J. Chaitin. *The limits of mathematics: a course on information theory & limits of formal reasoning.* Springer, New York, 1997.

[80] D. Challet and Y-C. Zhang. On the minority game: analytical and numerical studies. *Physica A*, 256(514):check this!, 1998.

[81] P. Checkland. *Systems thinking, systems practice.* John Wiley and Sons, Chichester UK, 1981.

[82] P. S. Churchland, Ramachandran V. S., and T. J. Sejnowski. A Critique of Pure Vision. In C. Koch and J. L. Davis, editors, *Large-Scale Neuronal Theories of the Brain.* A Bradford Book, The MIT Press, Cambridge, USA, 1994.

[83] C. W. Churchman. *Challenge to Reason.* McGraw-Hill, New York, USA, 1968.

[84] J. A. Clack. From Fins to Fingers. *Science*, 304:57–58, 2004.

[85] C. von Clausewitz. *On War.* transl. M. Howard and P. Paret, Princeton University Press, Princeton, USA, 1984.

[86] D. R. Conner. *Leading at the Edge of Chaos: How to Create the Nimble Organization* . Wiley, 1998.

[87] J. Copeland and R. Sylvan. Beyond the Universal Turing Machine. *Australasian Journal of Philosophy*, 77:46–66, 1999.

[88] P. A. Corning. The Re-emergence of "Emergence": A Venerable Concept in Search of a Theory. *Complexity*, 7(6):18–30, 2002.

[89] L. Correia. Self-organisation: a case for embodiment. In *Proceedings of The Evolution of Complexity Workshop at Artificial Life X: The 10th International Conference on the Simulation and Synthesis of Living Systems*, pages 111–116, 2006.

[90] J. M. Cortes, J. Marro, and J. J. Torres. Control of neural chaos by synaptic noise. *Biosystems*, in press, 2006.

[91] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.

[92] J. Crutchfield. The Calculi of Emergence: Computation, Dynamics, and Induction. *Physica D*, 75:11–54, 1994.

[93] J. P. Crutchfield. Is Anything Ever New? Considering Emergence. In G. Cowan, D. Pines, and D. Melzner, editors, *Complexity: Metaphors, Models, and Reality*, Santa Fe Institute Series in the Sciences of Complexity XIX, pages 479–497. Addison-Wesley, Redwood City, 1994.

[94] J. P. Crutchfield and D. P. Feldman. Statistical complexity of simple one-dimensional spin systems. *Phys. Rev. E*, 55(2):1239R1243R, 1997.

[95] J. P. Crutchfield and D. P. Feldman. Regularities Unseen, Randomness Observed: Levels of Entropy Convergence. *Chaos*, 13(1):25–54, 2003.

[96] J. P. Crutchfield and N. H. Packard. Symbolic Dynamics of Noisy Chaos. *Physica D*, 7(1-3):201–223, 1983.

[97] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108, 1989.

[98] R. Cummins. *Meaning and Mental Representation.* The MIT Press, Cambridge, USA, 1989.

[99] V. Darley. Emergent Phenomena and Complexity. *Alife IV*, 1994.

[100] P. C. W. Davies. Emergent biological principles and the computational properties of the universe. *Complexity*, 10(2):11–15, 2004.

[101] R. Dawkins. *The Selfish Gene.* Oxford University Press, Oxford, UK, 1976.

[102] R. Dawkins. Universal Darwinism. In D.S. Bendall, editor, *Evolution from Molecules to Men.* Cambridge University Press, 1983.

[103] T De Wolf and T. Holvoet. Emergence versus self-organisation: Different concepts but promising when combined. In S. Brueckner, G. D. M. Serugendo, A. Karageorgos, and R. Nagpal, editors, *Engineering Self-Organising Systems*, page 115. Springer, 2005.

[104] D. C. Dennett. *Brainstorms: Philosophical Essays on Mind and Psychology.* MIT Press, Cambridge, USA, 1978.

[105] D. C. Dennett. *Consciousness explained.* Little Brown and Co., Boston, USA, 1991.

[106] D. C. Dennett. Real Patterns. *Journal of Philosophy*, 88:27–51, 1991.

[107] R. Descartes. *A Discourse on Method: Meditations and Principles.* transl. J. Veitch, J. M. Dent and Sons, London, 1912.

[108] R. Descartes. *Discourse on the Method of Rightly Conducting the Reason, and Seeking Truth in the Sciences.* transl. L. Lafleur, Prentice Hall, originally published 1637, Englewood Cliffs, USA, 1960.

[109] J. Diamond. *Guns, Germs, and Steel: The Fates of Human Societies.* W.W. Norton & Company, New York, USA, 1999.

[110] J. Digby. Operations Research and Systems Analysis at RAND, 1948-1967. Technical Report N-2936-RC, The RAND Corporation, 1989.

[111] J. L. Ditty, S. B. Williams, and S. S. Golden. A Cyanobacterial Circadian Timing Mechanism. *Annual Review of Genetics*, 37(1):513–543, 2003.

[112] M. Dorigo, V. Maniezzo, and A. Colorni. Ant System: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 26(1):29–41, 1996.

[113] N. Dubilier, C. Mulders, T. Ferdelman, D. de Beer, A. Pernthaler, M. Klein, M. Wagner, C. Erseus, F. Thiermann, J. Krieger, O. Giere, and R. Amann. Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature*, 411(6835):298–302, 2001.

[114] P. Dudley, editor. *Bogdanov's Tektology, Book 1.* transl. V. N. Sadovsky, A. Kartashov, V. Kalle and P. Bystrov, Centre for Systems Studies, originally published 1910-1913, Hull, UK, 1996.

[115] W. Ebeling. Prediction and entropy of nonlinear dynamical systems and symbolic sequences with LRO. *Physica D*, 109:4252, 1997.

[116] B. Edmonds. Gossip, sexual recombination and the El Farol bar: modelling the emergence of heterogeneity. *Journal of Artificial Societies and Social Simulation*, 2(3):1–21, 1999.

[117] S. Eilon. The role of management science. *Journal of the Operational Research Society*, 31:17–28, 1980.

[118] P. Erdos and A. Renyi. On the strength of connectedness of random graphs. *Acta Mathematica Scientia Hungary*, 12:261–267, 1961.

[119] K-E. Eriksson and K. Lindgren. Structural information in self-organizing systems. *Physica Scripta*, 35:38897, 1987.

[120] M. D. Fagen, editor. *A History of Engineering and Science in the Bell System: National Service in War and Peace (1925-1975).* Bell Telephone Laboratories, Inc., 1978.

[121] J. Farey. *A Treatise on the Steam Engine: Historical, Practical, and Descriptive.* Longman, Rees, Orme, Brown, and Green, London, UK, 1827.

[122] D. P. Feldman and J. P. Crutchfield. Discovering Noncritical Organization: Statistical Mechanical, Information Theoretic, and Computational Views of Patterns in One-Dimensional Spin Systems. Technical Report 98-04-026, SFI Working Paper, 1998.

[123] D. P. Feldman and J. P. Crutchfield. Structural Information in Two-Dimensional Patterns: Entropy Convergence and Excess Entropy. *Physical Review E*, 67, 2003.

[124] S. Fish. Theory's Hope. *Critical Inquiry*, 30:374–378, 2004.

[125] R. L. Flood and M. C. Jackson. Total Systems Intervention: A Practical Face to Critical Systems Thinking. *Systems Practice*, 4(3):197–213, 1991.

[126] J. A. Fodor and Z. Pylyshyn. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28:3–71, 1988.

[127] J. W. Forrester. *Industrial Dynamics.* Pegasus Communications, Waltham, USA, 1961.

[128] M. Fortun and S. S. Schweber. Scientists and the Legacy of World War II: The Case of Operations Research (OR). *Social Studies of Science*, 23(4):595–642, 1993.

[129] J. F. C. Fuller. *The Foundations of the Science of War.* Hutchinson and Co. Ltd., London, 1925.

[130] M. Gardner. Mathematical Games: The Fantastic Combinations of John Conway's New Solitaire Game Life. *Scientific American*, 223:120–123, 1970.

[131] J. J. Garstka. Network Centric Warfare: An Overview of Emerging Theory. *Phalanx*, 33(4):1–33, 2000.

[132] M. Gell-Man. *The Quark and the Jaguar: Adventures in the Simple and the Complex.* Henry Holt and Company, New York, USA, 1994.

[133] M. Gell-Man and S. Lloyd. Effective Complexity. *Santa Fe Institute Working Paper*, 03-12-068, 2003.

[134] J. Giles. Stephen Wolfram: What kind of science is this? *Nature*, 417:216–218, 2002.

[135] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, Reading, UK, 1989.

[136] J. Goldstein. The Singular Nature of Emergent Levels: Suggestions for a Theory of Emergence. *Nonlinear Dynamics, Psychology, and Life Sciences*, 6(4), 2002.

[137] P. Grassberger. Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics*, 25:907938, 1986.

[138] A. Grisogono. Co-Adaptation. In *SPIE Symposium on Microelectronics, MEMS and Nanotechnology*, volume Paper 6039-1, Brisbane, Australia, 2005.

[139] A. Grisogono and A. J. Ryan. Designing Complex Adaptive Systems for Defence. In S. Wilson, editor, *System Engineering Test and Evaluation Conference*, Canberra, Australia, 2003.

[140] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* Springer-Verlag, New York, USA, 1983.

[141] J. Habermas. *Knowledge and Human Interests.* Heinemann, London, UK, 1972.

[142] H. Haken. *Synergetics: an introduction.* Springer, Berlin, Germany, 1983.

[143] A. D. Hall. *A Methodology for Systems Engineering.* Van Nostrand, Princeton, USA, 1962.

[144] J. Halley and D. Winkler. Towards consistent concepts of self-organization and self-assembly, in prep. 2006.

[145] G. Hardin. The Tragedy of the Commons. *Science*, 162(3859):1243–1248, 1968.

[146] G. P. Harmer and D. Abbott. Losing strategies can win by Parrondo's paradox. *Nature*, 402:864, 1999.

[147] V. Harte. *Plato on Parts and Wholes: The Metaphysics of Structure.* Oxford University Press, Oxford, 2002.

[148] C. Haskins, editor. *Systems Engineering Handbook: A guide for system life cycle processes and activities.* INCOSE, 2006.

[149] J. Haugeland. *Artificial Intelligence: The Very Idea.* MIT Press, Cambridge, USA, 1985.

[150] F. Heylighen. Modelling Emergence, World Futures. *Journal of General Evolution, special issue on creative evolution*, check this!, 1991.

[151] T. Hobbes. *Leviathan.* University of Oregon, Oregon, USA, 1651.

[152] D. R. Hofstadter. *Gödel, Escher, Bach: An eternal golden braid.* Vintage Books, New York, 1989.

[153] J. H. Holland. *Hidden Order: How Adaptation Builds Complexity.* Helix Books. AddisonWesley, Reading, USA, 1995.

[154] J. H. Holland. *Emergence : from chaos to order.* Helix Books. Perseus Books, Cambridge, USA, 1998.

[155] C. G. Holt and E. B. Dent. CAS in War, Bureaucratic Machine in Peace: The U.S. Air Force Example. *Emergence: The Journal of Complexity in Management and Organizations*, 3(3):90–107, 2001.

[156] J. Hoopes, editor. *Peirce on Signs: Writings on Semiotic by Charles Sanders Peirce*. University of North Carolina Press, Chapel Hill, USA, 1991.

[157] I. R. Hoos. *Systems analysis in social policy: a critical review*. Institute of Economic Affairs, London, UK, 1969.

[158] I. R. Hoos. Systems Techniques for Managing Society: A Critique. *Public Administration Review*, 33(2):157–164, 1973.

[159] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA*, 79(8):2554–2558, 1982.

[160] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. 70408.

[161] S. P. Hubbell, L. K. Johnson, E. Stanislav, B. Wilson, and H. Fowler. Foraging by bucket-brigade in leafcutter ants. *Biotropica*, 12(3):210213, 1980.

[162] C. Hull. *Principles of Behavior*. Appleton-Century-Crofts, New York, USA, 1943.

[163] D. Hume. *A Treatise of Human Nature*. Reprint Oxford University Press, 1978, Oxford, UK, 1741.

[164] D. Hume. *An Enquiry Concerning Human Understanding*. Ed. P. H. Niditch, Reprint Clarendon Press, 1975, Oxford, UK, 1777.

[165] A. Ilachinski. Land Warfare and Complexity, Part I: Mathematical Background and Technical Sourcebook. Technical report, Center for Naval Analyses, 1996.

[166] A. Ilachinski. Land Warfare and Complexity, Part II: An Assessment of the Applicability of Nonlinear Dynamics and Complex Systems Theory to the Study of Land Warfare. Technical report, Center for Naval Analyses, 1996.

[167] A. Ilachinski. *Artificial war : multiagent-based simulation of combat*. World Scientific Publishing, River Edge, New Jersey, USA, 2004.

[168] M. C. Jackson. Social systems theory and practice: the need for a critical approach. *International Journal of General Systems*, 10:135–151, 1985.

[169] M. C. Jackson. The Origins and Nature of Critical Systems Thinking. *Systems Practice*, 4(2):131–149, 1991.

[170] W. James. *A Pluralistic Universe*. University of Nebraska Press, originally published 1909, Lincoln, USA, 1996.

[171] N. Johnson, S. Rasmussen, C. Joslyn, L. Rocha, S. Smith, and M. Kantor. Symbiotic Intelligence: Self-organizing Knowledge on Distributed Networks Driven by Human Interaction . In *6th International Conference on Artificial Life*, University of California, California, USA, 1998.

[172] L. P. Kadanoff. Wolfram on Cellular Automata; A Clear and Very Personal Exposition. *Physics Today*, 55(7):55, 2002.

[173] I. Kant. *Critique of Pure Reason.* transl. N. Kemp Smith, Palgrave Macmillan, this edition published 1929, London, UK, 1781.

[174] S. A. Kauffman. Antichaos and Adaptation. *Scientific American*, 265:78–84, 1991.

[175] S. A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, Oxford, 1993.

[176] S. A. Kauffman. *Investigations.* Oxford University Press, Oxford, 2000.

[177] R.W. Kedzior. Evolution and Endurance: The U.S. Army Division in the Twentieth Century. Technical report, The Rand Corporation, 2000.

[178] M. J. Kelly. The Bell Telephone Laboratories - An Example of an Institute of Creative Technology. *Proceedings of the Royal Society of London A*, 203(1074):287–301, 1950.

[179] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, pages 1942–1948, Piscataway, USA, 1995.

[180] C. Kiefer and E. Joos. Decoherence: Concepts and Examples. *arXiv:quant-ph/9803052*, 1998.

[181] LTCOL D. Kilcullen. *Complex Warfighting.* Commonwealth of Australia, 2004.

[182] M. Kinnaird, L. Early, and B. Schofield. Defence Procurement Review 2003. Technical report, Department of the Prime Minister and Cabinet, 2003.

[183] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295:1662–1664, 2002.

[184] J. T. Klein. *Interdisciplinarity: History, Theory & Practice.* Wayne State University Press, Detroit, USA, 1989.

[185] S. J. Kline. *Conceptual Foundations for Multidisciplinary Thinking.* Stanford University Press, California, USA, 1995.

[186] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Organization of the information flow in the perception-action loop of evolved agents. In *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, page 177180. IEEE Computer Society, 2004.

[187] A. S. Klyubin, D. Polani, and C. L. Nehaniv. All else being equal be empowered. In M. S. Capcarr'ere, A. A. Freitas, P. J. Bentley, C. G. Johnson, and J. Timmis, editors, *Advances in Artificial Life, 8th European Conference, ECAL 2005*, volume 3630 of LNCS, page 744753. Springer, 2005.

[188] A. Korzibski. *Science and Sanity.* The International Non-Aristotelian Library, 1948.

[189] S. G. Krantz. Review: A new kind of science. *Bulletin of the American Mathematical Society*, 40(1):143–150, 2002.

[190] C. C. Krulak. The strategic corporal: Leadership in the Three Block War. *Marine Corps Gazette*, 83(1):18–22, 1999.

[191] R. Kurzweil. Reflections on Stephen Wolfram's "A New Kind of Science". `http://kurzweilai.net/articles/art0464.html?printable=1`, accessed 20/12/2006.

[192] F. W. Lanchester. *Aircraft in Warfare: The dawn of the Fourth Arm*. Constable and Co., London, UK, 1916.

[193] L. Lancieri. *Memory and forgetting, two complementary mechanisms to characterize the various actors of the Internet and their interactions*. Phd thesis, University of Caen, 2000.

[194] L. Lancieri. Reusing implicit cooperation, a novel approach in knowledge management. *TripleC (Cognition, Cooperation, Communication) international journal*, 2(1):28–46, 2004.

[195] B. Langefors. *Essays in Infology*. Studentlitteratur, Lund, 1995.

[196] C. Langton. Computation at the Edge of Chaos: Phase Transitions and Emergent Computation. In S. Forest, editor, *Emergent Computation*. MIT, 1991.

[197] J. S. Lansing. Complex Adaptive Systems. *Annual Review of Anthropology*, 32:183–204, 2003.

[198] P. S. Laplace. *Philosophical Essays on Probabilities*. transl. A. I. Dale, Springer-Verlag, originally published 1814, New York, USA, 1995.

[199] M. K. Lauren. Modelling Combat Using Fractals and the Statistics of Scaling Systems. *Military Operations Research*, 5(3):47–58, 2000.

[200] M. K. Lauren. A New Framework for Combat Risk Assessment Using the MANA Model. In *ORSNZ Conference*, Christchurch, NZ, 2001.

[201] R. Lee. Personal communication, 2004.

[202] G. W. Leibniz. *Dissertatio de Arte Combinatoria*. Leipzig, Germany, 1666.

[203] G. H. Lewes. *Problems of Life and Mind Vol 2*. Kegan Paul, Trench, Turbner, & Co., London, 1875.

[204] R. Lewin. *Complexity: Life at the Edge of Chaos*. Macmillan Publishing Co., New York, USA, 1992.

[205] R. C. Lewontin. The units of selection. *Annual Review of Ecology and Systematics*, 1:1–18, 1970.

[206] M. Li and P.M.B. Vitanyi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, New York, 2nd edition, 1997.

[207] W. Li. On the relationship between complexity and entropy for Markov chains and regular languages. *Complex Systems*, 5(4):381399, 1991.

[208] R. Lilienfeld. *The Rise of Systems Theory: An Ideological Analysis.* John Wiley & Sons, New York, USA, 1978.

[209] K. L. Lim, I. Mann, R. Santos, B. Tobin, M. J. Berryman, A. J. Ryan, and D. Abbott. Adaptive Battle Agents: Complex Adaptive Combat Models. *Proceedings SPIE*, 6039:48–60, 2005.

[210] K. Lindgren and M. G. Norhdal. Complexity measures and cellular automata. *Complex Systems*, 2(4):409440, 1988.

[211] A. S. Lipatov, E. A. Govorkova, R. J. Webby, H. Ozaki, M. Peiris, Y. Guan, L. Poon, and R. G. Webster. Influenza: Emergence and Control. *Journal of Virology*, 78(17):8951–8959, 2004.

[212] Lone Sentry. A Tactical Study of The Effectiveness of the German 88 MM Anti-Aircraft Gun As An Anti-Tank Weapon in the Libyan Battle. `http:www.lonesentry.com/tacticalstudy/index.html`, accessed 3/4/2006.

[213] C. Lucas. Complexity philosophy as a computing paradigm. In *Self-Organising Systems - Future Prospects for Computing Workshop*, UMIST, 1999.

[214] S. Luke, C. Cioffi-Revilla, L. Panait, and K. Sullivan. MASON: A New Multi-Agent Simulation Toolkit. In *Proceedings of the 2004 SwarmFest Workshop*, University of Michigan, 2004.

[215] M. W. Maier. Architecting principles for systems-of-systems. *Systems Engineering*, 1(4):267–284, 1998.

[216] D. Matthews. *Rethinking Systems Thinking: Towards a Postmodern Understanding of the Nature of Systemic Inquiry.* PhD thesis, University of Adelaide, 2004.

[217] H. Maturana and F. Varela. *The Tree of Knowledge: The Biological Roots of Human Understanding.* transl. J. Young, Shambhala Publications, originally published 1988, Boston, USA, 1984.

[218] J. McClelland and D. Rumelhart, editors. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition.* MIT Press, Cambridge, USA, 1986.

[219] W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115 – 133, 1943.

[220] B. McLaughlin and K. Bennett. Supervenience. In E.N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* The Metaphysics Research Lab, 2006.

[221] U. Merry. *Coping with Uncertainty : Insights from the New Sciences of Chaos, Self-Organization, and Complexity.* Praeger, 1995.

[222] G. Midgley, editor. *Systems thinking.* Sage library in business and management. Sage, London, 2003.

[223] J. S. Mill. *System of Logic.* Longmans, Green, Reader, and Dyer, 1843.

[224] J. Mingers. Recent Developments in Critical Management Science. *The Journal of the Operational Research Society*, 43(1):1–10, 1992.

[225] J. Mingers and S. Taylor. The Use of Soft Systems Methodology in Practice. *The Journal of the Operational Research Society*, 43(4):321–332, 1992.

[226] M. Mitchell, P. T. Hraber, and J. P. Crutchfield. Revisiting the edge of chaos: evolving cellular automata to perform computations. *Complex Systems*, 7:89139, 1993.

[227] J. Moffat. *Complexity Theory and Network Centric Warfare*. CCRP, Washington, USA, 2003.

[228] H. Moon. Looking forward and looking back: Integrating completion and sunk-cost effects within an escalation-of-commitment progress decision. *Journal of Applied Psychology*, 86:104–113, 2001.

[229] C. L. Morgan. *Emergent Evolution*. Williams and Norgate, London, 1923.

[230] C. L. Morgan. *Life, Mind and Spirit*. Williams and Norgate, London, 1926.

[231] C. L. Morgan. *The Emergence of Novelty*. Henry Holt and Co., New York, 1933.

[232] P. M. Morse and G. E. Kimball. *Methods of Operations Research*. John Wiley & Sons, New York, 1951.

[233] M. E. J. Newman. Assortative Mixing in Networks. *Phys. Rev. Lett.*, 89(208701), 2002.

[234] I. Newton. *Mathematical Principles of Natural Philosophy*. transl. B. Cohen & A. Whitman, University of California Press, this edition published 1999, Berkeley, USA, 1687.

[235] G. Nicolis and I. Prigogine. *Self-organization in nonequilibrium systems*. Wiley, New York, 1977.

[236] D. O. Norman and M. L. Kuras. Engineering Complex Systems. Technical report, MITRE, 2004.

[237] G. O'Brien. Connectionism, Analogicity and Mental Content. *Acta Analytica*, 22:111–131, 1998.

[238] G. O'Brien and J. Opie. Notes Toward a Structuralist Theory of Mental Representation. In H. Clapin, P. Staines, and P. Slezak, editors, *Representation in Mind: New Approaches to Mental Representation*. Elsevier Science, 2004.

[239] G. O'Brien and J. Opie. How do connectionist networks compute? *Cognitive Processing*, 7(1):30–41, 2006.

[240] T. O'Conner and H. Y. Wong. Emergent Properties. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, 2005.

[241] J. Opie. Personal communication, 2006.

[242] AIRCDRE K. Osley and C. Durrant. Complex Adaptive Systems Problems. In *2nd Complex Adaptive Systems for Defence Workshop*, University of Canberra, Canberra, Australia, 2005.

[243] S. E. Palmer. *Vision Science: Photons to Phenomenology.* MIT Press, Cambridge, USA, 1999.

[244] H. V. D. Parunak, S. Brueckner, M. Fleischer, and J. Odell. Co-X: Defining what Agents Do Together. In *Proceedings of Workshop on Teamwork and Coalition Formation.* AAMAS, 2002.

[245] R. Penrose. *Shadows of the Mind.* Oxford University Press, Oxford, UK, 1994.

[246] C. Perrow. *Normal Accidents: Living With High-Risk Technologies.* Basic Books, Inc., New York, USA, 1984.

[247] S. E. Phelan. A Note on the Correspondence Between Complexity and Systems Theory. *Systemic Practice and Action Research*, 12(3):237–245, 1999.

[248] D. Pitt. Mental Representation. In E.N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, 2005.

[249] M. Polanyi. *Personal Knowledge: Towards a Post-Critical Philosophy.* Routledge, London, UK, 1962.

[250] K. R. Popper. *Objective Knowledge.* Clarendon Press, Oxford, UK, 1972.

[251] K. R. Popper. Three Worlds. In *The Tanner Lecture on Human Values*, The University of Michigan, 1978.

[252] M. Prokopenko, F. Boschetti, and A. J. Ryan. An Information-theoretic Primer on Complexity, Self-organisation and Emergence. *Submitted to Complexity*, 2007.

[253] M. Prokopenko, V. Gerasimov, and I. Tanev. Evolving Spatiotemporal Coordination in a Modular Robotic System. In *Proceedings of the 9th International Conference on the Simulation of Adaptive Behavior (SAB'06)*, page 558569, Rome, Italy, 2006.

[254] M. Prokopenko, V. Gerasimov, and I. Tanev. Measuring Spatiotemporal Coordination in a Modular Robotic System. In L.M. Rocha, L.S. Yaeger, M.A. Bedau, D. Floreano, R.L. Goldstone, and A. Vespignani, editors, *Artificial Life X: Proceedings of The 10th International Conference on the Simulation and Synthesis of Living Systems*, pages 185–191, Bloomington IN, USA, 2006.

[255] M. Prokopenko, D. Polani, and P. Wang. Optimizing Potential Information Transfer with Self-referential Memory. In *Proceedings of 5th International Conference on Unconventional Computation (UC'06)*, pages 228–242, York, UK, 2006.

[256] M. Prokopenko, P. Wang, M. Foreman, P. Valencia, D. C. Price, and G. T. Poulton. On Connectivity of Reconfigurable Impact Networks in Ageless

Aerospace Vehicles. *Journal of Robotics and Autonomous Systems*, 53(1):36–58, 2005.

[257] M. Prokopenko, P. Wang, D. C. Price, P. Valencia, M. Foreman, and Farmer. A. J. Self-organizing Hierarchies in Sensor and Communication Networks. *Artificial Life, Special Issue on Dynamic Hierarchies,*, 11(4):407–426, 2005.

[258] A. Rapoport. Ludwig von Bertalanffy Memorial Lecture. In *S.G.S.R. annual meeting*, Boston, USA, 1976.

[259] A. Rapoport and A. M. Chammah. *Prisoner's Dilemma: A Study in Conflict and Cooperation*. University of Michigan Press, Ann Arbor, 1965.

[260] H. Reichenbach. *The Theory of Relativity and A-Priori Knowledge*. University of California Press, Berkeley, USA, 1920.

[261] C. Reynolds. Flocks, Herds, and Schools: A Distributed Behavioral Model. *Computer Graphics*, 21(4), July 1987.

[262] K. A. Richardson, G. Mathieson, and P. Cilliers. The Theory and Practice of Complexity Science: Epistemological Considerations for Military Operational Analysis. Technical report, Defence Science and Technology Laboratory, 2000.

[263] R. Rorty. *Philosophy and the mirror of nature*. Princeton University Press, Princeton, 1979.

[264] R. Rorty. Representation, social practise, and truth. In *Objectivity, Relativism and Truth: Philosophical Papers Volume 1*. Cambridge University Press, Cambridge, UK, 1991.

[265] R. Rosen. *Anticipatory Systems: Philosophical, Mathematical & Methodological Foundations*. Pergamon, New York, USA, 1985.

[266] Rutgers. The Virtual Biology Labs, Lab 1: Cell Structure. `http://bio.rutgers.edu/~gb101/virtuallabs_101.html`, accessed 12/6/2007.

[267] A. J. Ryan. About the bears and the bees: Adaptive approaches to asymmetric warfare. *Interjournal*, 2006.

[268] A. J. Ryan. Emergence is coupled to scope, not level. *Submitted to Complexity*, 2007.

[269] A. J. Ryan and A. Grisogono. Hybrid Complex Adaptive Engineered Systems: A Case Study in Defence. *Interjournal*, 2004.

[270] A. J. Ryan and D. O. Norman. Agile and Flexible Capability Development. In *Proceedings of the Land Warfare Conference*, Brisbane, Australia, 2006.

[271] A. J. Ryan, S. Wheeler, and T. Turner. Impact of Correlations in Logistics Supply Chains. In L. Caccetta and V. Rehbock, editors, *Proceedings ASOR Conference*, pages 146–153, Curtin University of Technology, Perth, Australia, 2005.

[272] S. N. Salthe. Summary of the Principles of Hierarchy Theory. `http://www.harmeny.com/twiki/pub/Main/SaltheResearchOnline/HT_principles.pdf`, accessed 10/6/2006.

[273] T. I. Sanders. *Strategic Thinking and the New Science: Planning in the Midst of Chaos Complexity and Change*. Free Press, 1998.

[274] J. P. Sartre. *Search for a Method*. trans. H. E. Barnes. Vintage, New York, 1968.

[275] F. de Saussure. *Course in General Linguistics*. transl. W. Baskin, The Philosophical Library, originally published 1916, New York, USA, 1959.

[276] J. J. Schneider. Black Lights: Chaos, Complexity, and the Promise of Information Warfare. *JFQ*, Spring, 1997.

[277] J. R. Searle. Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3:417–424, 1980.

[278] J. R. Searle. *Mind*. Oxford University Press, Oxford, UK, 2004.

[279] C. Shalizi. A New Kind of Science: A Rare Blend of Monster Raving Egomania and Utter Batshit Insanity. `http://www.cscs.umich.edu/~crshalizi/reviews/wolfram`, accessed 20/12/2006.

[280] C. R. Shalizi. *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata*. PhD thesis, University of Michigan, 2001.

[281] C. R. Shalizi. Methods and Techniques of Complex Systems Science: An Overview. *Arxiv:nlin.AO/0307015 v4*, 2006.

[282] C. R. Shalizi and J. P. Crutchfield. Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *Journal of Statistical Physics*, 104:819–881, 2001.

[283] C. R. Shalizi and K. L. Shalizi. Optimal Nonlinear Prediction of Random Fields on Networks. *Discrete Mathematics and Theoretical Computer Science*, AB(DMCS):11–30, 2003.

[284] C. R. Shalizi and K. L. Shalizi. Blind Construction of Optimal Nonlinear Recursive Predictors for Discrete Sequences. In M. Chickering and J. Joseph Halpern, editors, *Uncertainty in Artificial Intelligence: Proceedings of the Twentieth Conference*, pages 504–511, Arlington, Virginia, 2004. AUAI Press.

[285] C. R. Shalizi, K. L. Shalizi, and R. Haslinger. Quantifying self-organization with optimal predictors. *Physical Review Letters*, 93(11):11870114, 2004.

[286] A. Shamir. How to Share a Secret. *Communications of the ACM*, 22(11):612–613, 1979.

[287] C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October, 1948.

[288] J. A. Shapiro. Bacteria as Multicellular Organisms. *Scientific American*, 258(6):82–89, 1988.

[289] R. Shaw. *The Dripping Faucet as a Model Chaotic System*. Aerial Press, Santa Cruz, California, 1984.

[290] K. R. Jr Short. The Value of the Mathematical Form. `http://www.gwu.edu/~asc/biographies/ashby/MATRIX/V/v_1.html`, 2007.

[291] H. T. Siegelmann. *Neural Networks and Analog Computation: Beyond the Turing Limit*. Springer Verlag, 1999.

[292] H. A. Simon. The Architecture of Complexity. *Proceedings of the American Philosphical Society*, 106(6):1962, 1962.

[293] H. A. Simon. Can there be a science of complex systems? *Interjournal*, 509, 2000.

[294] F. Slanina. Social organization in the minority game model. *Physica A*, 286:367–376, 2000.

[295] E. A. Smith. *Effects Based Operations: Applying Network Centric Warfare In Peace, Crisis, And War*. CCRP Publication Series, 2002.

[296] R. V. Solé and S. Valverde. Information theory of complex networks: on evolution and architectural constraints. In E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, editors, *Complex Networks*, volume 650 of *Lecture Notes in Physics*. Springer, 2004.

[297] G. Sommerhoff. The Abstract Characteristics of Living Systems. In F. E. Emery, editor, *Systems Thinking: Selected Readings*. Penguin Books, Harmondsworth, UK, 1969.

[298] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, The MIT Press, Cambridge, 1998.

[299] Deutsches Elektronen Synchrotron. Glossary. `http://zms.desy.de/press/glossary/index_eng.html`, accessed 14/12/2006.

[300] V. L. Syrmos, C. T. Abdallah, P. Dorato, and K. Grigoriadis. Static Output Feedback: A Survey. *Automatica*, 33(2):125–137, 1997.

[301] A. M. Turing. On Computable Numbers, With An Application To The Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(2):230–265, 1936-7.

[302] W. Ulrich. An Appreciation of C. West Churchman. `http://www.geocities.com/csh_home/cwc_appreciation.html`, accessed 21/12/2006. first published as C. West Churchman - 75 years, Systems Practice, 1(4):341-350, 1988.

[303] W. Ulrich. *Critical Heuristics of Scoial Planning: A New Approach to Practical Philosophy*. Haupt, Berne, 1983.

[304] Van Gelder, T. What might cognition be, if not computation? *Journal of Philosophy*, XCII(7):345–381, 1995.

[305] R. Van Gulick. Reduction, Emergence and Other Recent Options on the Mind-Body Problem: A Philosophic Overview. *Journal of Consciousness Studies*, 8:1–34, 2001.

[306] D. P. Varn. *Language Extraction from ZnS.* Phd thesis, University of Tennessee, 2001.

[307] L. von Bertalanffy. *Modern theories of development: An introduction to theoretical biology.* transl. J. H. Woodger, Oxford University Press, originally published 1928, New York, USA, 1933.

[308] L. von Bertalanffy. The Theory of Open Systems in Physics and Biology. *Science*, 111(2872):23–29, 1950.

[309] L. von Bertalanffy. General System Theory. *General Systems*, 1:1–10, 1956.

[310] L. von Bertalanffy. *General System Theory: Foundations, Development, Applications.* George Brazilier, New York, USA, 1969.

[311] B. Von Eckardt. On Evaluating the Scientific Status of Psychoanalysis. *The Journal of Philosophy*, 78(10):570–572, 1981.

[312] B. Von Eckardt. *What Is Cognitive Science?* A Bradford book, The MIT Press, Cambridge, USA, 1993.

[313] H. von Foerster. Perception of the Future and the Future of Perception. *Instructional Science*, 1(1):31–43, 1972.

[314] H. von Foerster. Cybernetics of Cybernetics. In K. Krippendork, editor, *Communication and Control in Society.* Gordon and Breach, New York, USA, 1979.

[315] H. von Foerster. Ethics and Second Order Cybernetics. *Cybernetics and Human Knowing*, 1(1):9–19, 1992.

[316] J. von Neumann. *Theory of Self-Reproducing Automata.* Burks, A. W., editor, University of Illonois Press, Urbana, USA, 1966.

[317] A. Wagner. *Robustness and Evolvability in Living Systems.* Princeton University Press, Princeton, USA, 2005.

[318] LTCOL D. Wainwright. *Adaptive Campaigning: The Land Force Response to Complex Warfighting.* Commonwealth of Australia, 2006.

[319] M. M. Waldrop. *Complexity: The Emerging Science at the Edge of Order and Chaos.* Simon & Schuster, New York, USA, 1992.

[320] J. Walsh. Bell Labs: A Systems Approach to Innovation Is the Main Thing. *Science*, 153(3734):393–396, 1966.

[321] J. B. Watson. *Psychology: From the Standpoint of a Behaviorist.* J. B. Lippincott, Philadelphia, USA, 1919.

[322] D. J. Watts and S. H. Strogatz. Collective Dynamics of 'Small-World' Networks. *Nature*, 393:440–442, 1998.

[323] W. Weaver. Science and Complexity. *American Scientist*, 36(536), 1948.

[324] G. M. Weinberg. *An Introduction to General Systems Thinking.* Dorset House Publishing, New York, USA, Silver Anniversary edition, 2001.

[325] S. Wheeler, A. J. Ryan, and T. Turner. Modelling Logistics Chains for Optimal Supply and Maximum Throughput. Technical report, DSTO, 2006.

[326] N. Wiener. Cybernetics. *Scientific American*, pages 14–19, 1948.

[327] Wikipedia contributors. Formal system, accessed 22/1/2007.

[328] U. Wilensky. NetLogo, 1999.

[329] O. Williamson. Hierarchical Control and Optimal Firm Size. *Journal of Political Economy*, pages 123–138, 1967.

[330] E. O. Wilson. *Sociobiology: the new synthesis.* Belknap Press of Harvard University Press, Cambridge, USA, 1975.

[331] K. G. Wilson. Problems in Physics with Many Scales of Lengths. *Scientific American*, 241(2):158–179, 1979.

[332] S. Wolfram. Universality and Complexity in Cellular Automata. *Physica D*, 10, 1984.

[333] S. Wolfram. *A New Kind of Science.* Wolfram Media, Inc., Champaign, USA, 2002.

[334] D. H. Wolpert and K. Tumer. An Introduction To Collective Intelligence. Technical report, NASA Ames Research Laboratory, 1999.

[335] A. Wuensche. Classifying Cellular Automata Automatically: Finding gliders, filtering, and relating space-time patterns, attractor basins, and the Z parameter. *Complexity*, 4(3):47–66, 1999.

[336] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.