On the evolutionary constraints necessary for immunity

by

Thomas Edward Hebbron

Thesis for the degree of Master of Philosophy

August 2015
All organisms represent a tempting storehouse of resources for other organisms to exploit. This fact leads to the evolution of semi-permeable defensive boundaries. Nutrients must be allowed in, and waste products allowed out. Aggressors should be kept out if possible or identified and attacked once they breach the boundary. An ongoing co-evolutionary battle leads, on the defensive side, to the finely tuned mechanisms of the immune system.

The thesis takes a functional perspective on immune system evolution. We begin with a historical review of the scientific understanding of immunity, and then ask what features of the immune systems we see today are general enough to occur again if we replayed the tape of life on Earth, and what features are likely to be merely contingent. There are computationally useful processes found in immune systems, but separating these transcendental principles from observable biological implementations is difficult.

If pathogens had infinite flexibility to change their appearance, or if hosts had similar flexibility to alter their metabolism, the co-evolutionary battle between them could not occur. We therefore argue that an evolutionary constraint necessary for immune system evolution is the concept of entrenchment or ‘lock-in’. We use a variation on the NK-landscape model to show how the structure of the dependency network between traits affects both evolvability and entrenchment and thus promotes the evolution of immune systems.
Contents

Declaration of Authorship xi
Acknowledgements xiii

1 Introduction 1
  1.1 The city and the immune system . . . . . . . . . . . . . . . . . . . . . . . . . 3
    1.1.1 The city, reduced . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
    1.1.2 The evolution of the city . . . . . . . . . . . . . . . . . . . . . . . . . . 4
    1.1.3 What about immunology? . . . . . . . . . . . . . . . . . . . . . . . . . 5
  1.2 A functional perspective on immune evolution . . . . . . . . . . . . . . . . . . 5
  1.3 Organised complexity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
  1.4 Specific research questions . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
    1.4.1 RQ1: Is there a class of problem space where evolutionary lock-in can
    provide slower evolving hosts sufficient information to identify faster
    evolving pathogens? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
    1.4.2 RQ2: Do more ‘biologically-inspired’ epistatic graphs exhibit greater
    ‘lock-in’? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
    1.4.3 RQ3: Can a ‘biologically-inspired’ epistatic graph improve evolvability? 11
  1.5 Thesis structure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
  1.6 Publications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2 Background 13
  2.1 Immunology: a history of ideas . . . . . . . . . . . . . . . . . . . . . . . . . . 16
    2.1.1 Smallpox and inoculation . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
    2.1.2 Germ theory: the microbiological revolution . . . . . . . . . . . . . . . 17
    2.1.3 Immunisation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
    2.1.4 Cellular and humoral immunology . . . . . . . . . . . . . . . . . . . . . 22
  2.2 Modern immunology . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
    2.2.1 Adaptive immune system . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
    2.2.2 Innate immune system . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
    2.2.3 Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
  2.3 Biosemiotics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
  2.4 Evolutionary biology . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
    2.4.1 Evolutionary explanations . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
    2.4.2 Evolutionary history of immune systems . . . . . . . . . . . . . . . . . . 35
  2.5 Artificial Immune Systems . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
    2.5.1 The literal school . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
    2.5.2 The metaphorical school . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
    2.5.3 Computational immunology . . . . . . . . . . . . . . . . . . . . . . . . . . 41
## List of Figures

1. **Problems of simplicity, disorganised complexity and organised complexity.**  
   - Figure 1.1: Problems of simplicity, disorganised complexity and organised complexity.  
   - Pages: 8

2. **An outline conceptual framework for a bio-inspired computational domain. (Stepney et al., 2005)***  
   - Figure 2.1: An outline conceptual framework for a bio-inspired computational domain. (Stepney et al., 2005)  
   - Pages: 15

2. **Space of models that capture a desired property of interest. (Hart et al., 2013)***  
   - Figure 2.2: Space of models that capture a desired property of interest. (Hart et al., 2013)  
   - Pages: 15

2. **Phagocytosis***  
   - Figure 2.3: Phagocytosis  
   - Pages: 30

2. **Partitioning the universe of antigens***  
   - Figure 2.4: Partitioning the universe of antigens  
   - Pages: 32

3. **Calculating fitness contributions for an \(N = 3, K = 2\) fitness landscape example.***  
   - Figure 3.1: Calculating fitness contributions for an \(N = 3, K = 2\) fitness landscape example.  
   - Pages: 49

4. **Adjacency matrix and network diagram for each class of epistatic network.***  
   - Figure 4.1: Adjacency matrix and network diagram for each class of epistatic network.  
   - Pages: 59

4. **Untuned assortativity values for epistatic networks generated using the \(NK\alpha\) method \emph{before} any assortative edge-swapping is performed.***  
   - Figure 4.2: Untuned assortativity values for epistatic networks generated using the \(NK\alpha\) method \emph{before} any assortative edge-swapping is performed.  
   - Pages: 60

4. **Edge-swapping procedure to generate networks with desired assortativity \(r\).***  
   - Figure 4.3: Edge-swapping procedure to generate networks with desired assortativity \(r\).  
   - Pages: 61

4. **Assortative and dissortative epistatic maps for \(N = 32, K = 8\) and \(\alpha = 1.5\).***  
   - Figure 4.4: Assortative and dissortative epistatic maps for \(N = 32, K = 8\) and \(\alpha = 1.5\).  
   - Pages: 61

4. **The asymmetry of accessible assortative and dissortative configurations when rewiring graphs for selected \(r\).***  
   - Figure 4.5: The asymmetry of accessible assortative and dissortative configurations when rewiring graphs for selected \(r\).  
   - Pages: 62

4. **Length of adaptive walks under uniform, random, and scale-free epistasis.***  
   - Figure 4.6: Length of adaptive walks under uniform, random, and scale-free epistasis.  
   - Pages: 68

4. **Optimum fitness reached by adaptive walks under uniform, random, and scale-free epistasis.***  
   - Figure 4.7: Optimum fitness reached by adaptive walks under uniform, random, and scale-free epistasis.  
   - Pages: 69

4. **An assay of neighbour fitness at selected points on different fitness landscapes.***  
   - Figure 4.8: An assay of neighbour fitness at selected points on different fitness landscapes.  
   - Pages: 70

4. **Correlation between walk step number and fitness, number of mutations tested, and maximum mutant degree.***  
   - Figure 4.9: Correlation between walk step number and fitness, number of mutations tested, and maximum mutant degree.  
   - Pages: 71

4. **The 'Massif Central' global feature of \(NK\) landscapes.***  
   - Figure 4.10: The 'Massif Central' global feature of \(NK\) landscapes.  
   - Pages: 72

4. **The Massif Central is not significantly affected by assortativity.***  
   - Figure 4.11: The Massif Central is not significantly affected by assortativity.  
   - Pages: 73

4. **Basins of attraction on \(NK\) landscapes.***  
   - Figure 4.12: Basins of attraction on \(NK\) landscapes.  
   - Pages: 74

4. **Comparing the length of adaptive walks and fitness of optima on landscapes with assortative and dissortative graphs.***  
   - Figure 4.13: Comparing the length of adaptive walks and fitness of optima on landscapes with assortative and dissortative graphs.  
   - Pages: 75

4. **Conservation of individual alleles.***  
   - Figure 4.14: Conservation of individual alleles.  
   - Pages: 76

4. **Distributions of per-locus correlation between out-degree and predictive utility and out-degree and uniqueness.***  
   - Figure 4.15: Distributions of per-locus correlation between out-degree and predictive utility and out-degree and uniqueness.  
   - Pages: 77
List of Tables

4.1 Size $n$ and assortativity coefficient $r$ for various networks, taken from Newman (2002), see reference for details. ................................................................. 58
Declaration of Authorship

I, Thomas Edward Hebbron, declare that the thesis entitled *On the evolutionary constraints necessary for immunity* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: Hebbron et al. (2008), Hebbron et al. (2009) and Hebbron and Noble (2009)

Signed: ........................................................................................................................................

Date: ........................................................................................................................................
Acknowledgements

To my supervisors, Jason Noble and Seth Bullock, thank you for the freedom to pursue this line of research. Your initial caution at my interest in immune systems was well founded — there is so much to read and so little time!

Jason, as my principal supervisor, your endless patience and enthusiasm enabled me to keep going when things looked bleak. I wish you & Sirem the very best for your new life in the sun!

My interest in immune systems goes back to discussions with Sri Dasmahapatra during the Complexity Science MSc at Southampton, so some of the thanks, and blame, should lie with Sri for starting me out on this path.

Many others have contributed in some way to the development of the ideas presented here. In particular the members of the Science & Engineering of Natural Systems (SENSe) Group for the frequent discussions over coffee, lunch, or beer; these are much missed.

Finally, to Philippa, you have supported me throughout this effort, for all this time. Thank you for everything.
Introduction

This thesis is about the immune system, its function, and the problem it solves. Whether we are looking at the more familiar adaptive immune system in higher vertebrates, or the more ancient innate immune systems, found in plants, insects and animals, we can be sure that these sophisticated mechanisms have evolved under extreme selective pressure — the need to engage with the constant sea of microbiological agents in and around us. It’s not all war — we depend on our commensal microbiota: the gut bacteria that help digest our food, the benign species that displace pathogenic competitors on and in us\(^1\).

In any evolved artefact, we can make out some of the effects of the evolutionary history that shaped its present state. The scars of failure, the ‘design’ apparent in the ways it nestles in its niche. As the ‘blind watchmaker’ tried many possible variations and discarded the failures, gradually the bounds of the viable were revealed. Baked into the eventual artefact are the compromises of satisfying many constraints. Some arise from physical laws (e.g., thermodynamics), some from the geometry of the problem at hand (e.g., shaping a lens to focus an image), some from the materials available to work with (e.g., computing with neurons, not silicon), and some from the history that came before (e.g., the route the recurrent laryngeal nerve famously takes from a giraffe’s brain to its voice box, circuitously via its heart (Dawkins, 2009)). Reverse engineering

\(^1\)We are perhaps only actually 10% human. In a healthy person, microorganisms account for 1 to 3% of body mass, but outnumber human cells 10 to 1. The huge quantity of genetic material this introduces is actively being mapped (Huttenhower et al., 2012).
these complex ensembles to reveal why they operate the way that they do can be enormously difficult. But, as life has had a few billion years head-start, it is advantageous to truly understand the remarkable and robust solutions it has discovered — particularly to the problem of solving complex problems. Applying this reverse engineering process, to immune systems specifically, is the core subject of this thesis. Our contribution is in finding an identifying property of a class of problems that are amenable to immune type algorithms. Those in the Artificial Immune Systems field or those applying their work will benefit from this thesis when deciding whether an immune-inspired algorithm is appropriate to a new domain.

We begin with an analogy on a human scale. A city, like a cell, comes about because there are tangible benefits to operating in a controlled environment. Making your living out in the wild is dangerous and expensive: there are many risks and associated costs when your whole enterprise is exposed to any and all random disturbances. Much better to shut out the world; build a boundary and concentrate everything required for life inside it — a self-maintaining microcosm.

Concentrated inside the protective enclosure, economies of scale can take off — security and other costs are shared and the close proximity of people, with their wants and surpluses, encourages collaboration on the task of living: working in teams, forging new interactions and achieving productivity beyond basic survival. Individually freed from the general pressures of maintaining security or growing food, specialisation becomes possible. With the lower strata of needs (food and shelter) fulfilled by shared labour, higher strata of behaviour are free to develop. It becomes possible to excel at some niche but useful skill whose utility outside the boundary, without the support of this controlled and rich environment, would be unsustainable: all your time and energy would go on basic sustenance.

The enhanced productivity possible by collaborative effort creates a local surplus, an immediate lure for both outsiders and inhabitants, tempted to renege on the social contract, for individual profit. The existence of local surplus creates a niche for an exploitative strategy to inhabit. The existence of exploitation in turn drives the development of defences to protect the fruits of communal labour from bandits and cheats.

Constructing even the most rudimentary defensive boundary marks an important event in the transition from settlement to city: enclosing and fortifying a new entity, drawing the line between inside and out and further consolidating the shared identity within the walls.

No city is an island, entire of itself: interactions with the outside world are fundamental to its sustained existence. The ideal defensive wall is both there and not there, impenetrable when the city is under attack, but no barrier to trade and the normal required comings and goings of the city. Gates, occasionally punctuating the wall, solve this functional conflict: the wall is built as securely as needs be, while the gates are built to be the best compromise between security and selective accessibility.
Arbitrary interactions are prevented by the perimeter wall, the first line of defence, funnelling all traffic through the gates where it can be scrutinised by the guards, a second line of defence, to focus attention and differentiate between benign traders and recognised troublemakers. Patrols inside the city hunt for miscreants who manage to get in, inhabitants of the city who cheat the community, or, particularly dangerous, rogue guards. Known offenders are swiftly dealt with, whereas those previously unknown and detected by their offences are identified and added to the city’s blacklist. Offenders could be recognised by a number of features: face, clothes, weapons, or evidence on the person. The consequences of discovery by the guards will depend on the offence committed, the status of the offender and the current state of the city. The guards also maintain the integrity of the first line of defence: attacks on the wall prompt counter-attacks against assailants. If the outer wall is breached, guards arrive to defend the gap and recruit workers to repair it.

We take the city and its security as a metaphor for the immune system. The defence mechanisms of the city have metaphorical parallels in the defence mechanisms that organisms employ against microbiological threats: bacteria, viruses, fungi and parasites.

As the wall is to the city, the skin is to the human. The importance of the skin, the epidermis, the first layer of defence against microbiological assailants, is most apparent when it is breached: wounds and burns are a major cause of disease and death. The natural gateways to the body are lined with mucous membranes, creating a hostile environment for pathogens, and secretions such as saliva and tears contain active antibiotic substances. Inside the body a complex ensemble of cellular and molecular components correspond to the guards and their weapons: on patrol, recognising and dispatching pathogens, learning and remembering their characters for subsequent encounters.

Students of the history of the city face the problem of reconstructing the interacting pattern of forces that shaped the emergence of the earliest cities. Similarly, immunology must reconstruct the history of adaptive pressures and the resulting phylogenetic pathways that lead us to the modern immune system in order to fully understand it. However, as we will see in this chapter and the next, immunology has typically been confined to questions of mechanism rather than questions of evolutionary history. This thesis employs computational modelling techniques in an attempt to redress that balance.

1.1 The city and the immune system

We have already identified some raw similarities in the defence mechanisms of both cities and organisms, but we can take this metaphor further, to parallels in methodology: how can we understand the defences of the city or of the body?
1.1.1 The city, reduced

The organisation and purpose of the defences are most apparent when under attack. Guards are mobilised to bar the gates and protect the outer wall: preparing boiling water to pour on attackers through murder holes, firing bows from arrow loops and battlements. The attackers have, of course, their own strategies, coevolved with the development of fortifications. Their principal aim is to breach the perimeter: using battering rams to break down gates; catapults to smash the walls or sappers to undermine and collapse them from beneath, or simply to tunnel into the city. Strategies on both sides of the wall are subject to constant subversion and innovation. Tunnelling, for instance, was detected by watching for ripples in strategically placed cauldrons of water, and subverted by digging counter-tunnels then meeting the attackers with angry bees, smoke, or by flooding.

To understand the defences of the city, towards either improving or defeating them, we could take a reductionist approach: subdivide the city and study in great detail the construction and operation of the gates; an intruder; the commander of the guards: any component that appears to have a significant role in the operation of the whole. This approach will produce a detailed and hopefully predictive model for the particular city, but does not answer more general explanatory questions as to why the defences observed evolved as they have. Walled cities proved a dominant way of life for many centuries, arising in the geographically and socially unconnected civilisations of Mesoamerica and Europe. A biologist would see this as an example of convergent evolution. A truly explanatory account must detail the pressures that prompted the independent adoption of the same underlying strategy.

1.1.2 The evolution of the city

An evolutionary account of the city tells how, in the late Neolithic, as human populations grew and unexploited territory and game populations shrunk, man was forced into a change of lifestyle. Agriculture arose from a necessity to produce more from the land, and this investment into the land led to the establishment of more permanent human settlements (Mumford, 1968; Kostof and Tobias, 1999; Bairoch, 1991; O’Flaherty, 2005).

As settlements grew and agriculture became more efficient, economies of scale allowed the production of surplus food and goods. Concentration of surplus led to the emergence of the first cities: centres of population and production, connected by trade. The growing wealth of the city needed protecting from individual opportunists and marauding armies outside its boundaries, and inside, from cheats in the community. Again, the economy of scale allowed substantial defences to be erected in the common interest, and drove specialisation in the labour force: beyond farmers and tool makers, the community begins to need builders, soldiers, watchmen, and administrators
to coordinate all this activity. O’Flaherty (2005) argues that military defence is one of the most compelling forces that drove the emergence of the city.

1.1.3 What about immunology?

The anthropological account above explains city formation as a series of responses to pressures on the population over time. Common pressures prompt similar responses, and so we see the form of the walled city repeated.

But what about immunology? The field has generally taken a reductionist approach akin to the piece-by-piece analysis of the city described in Section 1.1.1. As components of the immune system were discovered, they were studied to determine their immediate role in defence. Once the mechanism of a particular immune component was understood, the business of assisting or damping in the service of some therapeutic gain could begin.

The early successes of inoculation and antibiotics made a huge impact on medicine. Obviously exogenous, bacterial and viral pathogens were progressively discovered and conquered. This led to the establishment of a self / non-self distinction as the paradigm for understanding the immune system: its primary goal was seen as the classification of any agent as being either part of the body or of external origin, enabling the appropriate response in each case.

Sadly, this clear dichotomy soon becomes blurred when we consider cancer, autoimmune diseases and immune-targeting diseases such as HIV/AIDS. These require a much deeper understanding of the complex system of components and interactions that make up the immune system if we are to enjoy similar therapeutic success. As with most complex systems, the properties of interest (e.g., the nebulous ‘self’) are emergent, and a reductionist methodology is not sufficient to understand the immune system with the fidelity required to conquer autoimmune and immune-targeting diseases (Vance, 2000; Bersini, 2005; Cohen, 2004).

The anthropological route taken to understand the emergence and progress of city defences illustrates the perspective we will take on the evolution of the immune system: looking for general evolutionary pressures that cause particular paths through design space to be taken. We will argue that it is only through this type of approach that a full understanding of the immune system and its pathologies can be reached.

1.2 A functional perspective on immune evolution

Gould (1990) argued that if it were possible to ‘replay the tape’ of evolution, the life that resulted, both in general and in detail, would be very different. The particular mechanics of, for example,
locomotion or sensory organs could certainly have been different. However, it seems harder to imagine that these functional roles would not be fulfilled by some sort of biological mechanism, that certain strategic mechanisms we observe transcend a particular instance: historical progressions, such as the history of life, are the product of both contingency and necessity (Fontana and Buss, 1994). This is an excellent way to get to the heart of our questions about immune systems: what features of their design and the countermeasures they inspire are general enough to appear in a hypothetical replay of the course of evolution? Conversely, what features of modern immune systems are merely contingent, the result of historical accident? There are certain features of living organisms, situated in an ecology, that seem highly likely to be universal: that an individual organism represents a bag of resources that provides an incentive for predation or parasitism is one such fact. At this basic level, such a claim is not controversial, but, how far up do the generalities go? In a hypothetical ecology similar to ours, containing creatures on the order of complexity of mammals, would immune system mechanisms corresponding to B and T lymphocytes (for example) always occur?

We argue that the science of immunology has, to date, neglected these kinds of functional questions in favour of an understandable emphasis on mechanisms. In other words, how does the immune system work, how might we fix it when it’s broken, and what is its normal course of development? The question of function (why) however, is often left implicit — immune systems are for protecting against pathogens. We believe that ‘why’ questions are at least as important as ‘how’ questions, that the mass of empirical data now available on the phylogeny and mechanism of real immune systems represents a significant opportunity, and that advances in simulation modelling techniques mean that it is now reasonable to ask questions about the history of selection that led to the modern immune system. We are not alone in this belief. In Tending Adam’s Garden, Cohen writes: “For the past century immunology has, with great success, been occupied with analysing the immune system into its molecular building-blocks. The field is now ripe for synthesis.” (Cohen, 2004, p. 30).

We are interested not just in the types of immune systems that might arise if you could replay the tape of life from the beginning using the same physico-chemical building blocks, but those that might arise from different building blocks altogether. This question may not yet be of interest to the biological immunology community, but the immune system is a source of inspiration in other fields where such a question is fundamental to their aims. We will see in Section 2.5 and Chapter 3 that much of the existing work on computational models of the immune system in fact consists of pragmatic borrowing of abstracted immune system principles in order to exploit them in an engineering context, a core activity in the field of Artificial Immune Systems (AIS) research. Care must be taken when abstracting and transferring strategies and designs in this way. If there are truly underlying principles that transcend the particular domain of immune (or city) defence and these can be abstracted and applied to congruent problem domains, then they must be extracted from the specific developmental history of a single instance; i.e., the research must
go beyond the contingent constraints and compromises particular to a single instance as these obscure the relevant design patterns.

In the same way that the building of city walls reinforces the individuality and the identity of the city and its population, having an immune system seems itself to be a marker of individuality in the biological world. Immune systems are difficult to discuss without referring to the self / non-self distinction, and yet in the absence of any immune response, as protection from ‘outside’, the boundary of an individual is difficult to distinguish at all. It therefore seems likely that the evolutionary history of individuality and the immune response are inextricably intertwined.

We propose the following set of major transitions in the history of immune system evolution. Here, the list is given solely to orient the reader to where our specific questions and contributions lie in a programme of research that extends beyond this thesis. The transitions are discussed further in Chapter 2:

1. The emergence of membranes; early individuation.

2. Eukaryogenesis: breaking free of the volume / surface area size constraint on early prokaryotic life.


4. Inducible innate defences: the emergence of transitory biochemical weapons and sensors to trigger their synthesis and release, supplementing expensive constitutive (ever-present) defences.


6. Emergence of acquired immunity: the origins of the sophisticated ‘sensor upgrade’ the immune systems of jawed vertebrates acquired approximately 500 million years ago. Increased specificity and immunological memory enables adaptation to pathogens within a single lifetime, rather than over evolutionary time as with the innate immune systems that all other living things have (Matsunaga and Rahman, 1998; Laird et al., 2000).

These levels also represent an increasing degree of complexity in the world, as all members of the pathogenic, parasitic or predatory ecosystem co-adapt. Often, such adaptation reflexively makes the world not just different, but more complex still. Information about the world, good strategies and corresponding counter-strategies, are accumulated in the genomes of those who succeed (see Godfrey-Smith, 1998, for a parallel argument regarding the evolution of cognitive systems).

In either borrowing from the biological immune system to inspire our own engineered systems, or to better understand the biological system in order to push its complex network of components into therapeutic states, we need to gain the ability to separate the contingent details of the specific implementation from the necessary, strategic elements.
1.3 Organised complexity

![Diagram showing problems of simplicity, disorganised complexity, and organised complexity]

Figure 1.1: Adapted from Lima (2011, Chapter 2)

Weaver (1948) neatly classified the kinds of problems that scientists are interested in into simple, disorganised complex, and organised complex systems, as illustrated in Figure 1.1.

**Simple** When there are no interactions between elements of a simple system, such as the (idealised) physics of firing a cannonball, we have a simple system, easily understood through linear equations.

**Disorganised complex** If we have many interactions, but no discernible structure to those interactions, we have a disorganised complex system: for example the gas laws and other particle systems described by statistical physics.

**Organised complex** Finally, in the ‘interesting in-between’, organised complexity is the result of structured interactions between specific parts of the system. In these systems, structure contributes massively to function. This structure is not random, it is appropriate to the task; i.e., it incorporates information about the task. In biological systems, the structure has, of course, come about through a history of selection.

The structures we find in these naturally evolved ‘organised complex’ networks are of huge interest to scientists and engineers. They represent a several-billion-year head start, using a massively parallel search process, on the problem of solving complex problems. Investigating the role that network topology plays in underpinning immune system dynamics will be a key part of this thesis.

1.4 Specific research questions

Immune defences are of fundamental importance to biological life, and must therefore have influenced, and been influenced by, the networks that govern biological evolution and development.
Immune systems solve a certain class of problem — recognising biological agents that often evolve faster than the host. We propose that by examining network properties, that promote and constrain evolvability, we will gain insight into the bounds of this niche, where immune algorithms succeed.

1.4.1 RQ1: Is there a class of problem space where evolutionary lock-in can provide slower evolving hosts sufficient information to identify faster evolving pathogens?

What would have to change about the world for immune systems to cease to exist? The basic story of immunity is a long term co-evolutionary arms race between the immune system sensors and evasive adaptations of pathogens. What if pathogens had an infinite variety of coats to wear, and the cost of changing them was zero? The immune system would be effectively blinded, unable to perform recognition. Immune responses are damaging to the host at the best of times, with recognition helping to ensure they are targeted at the pathogens. Without recognition, the collateral damage would be severe, and indeed, if the only signal that initiates a response comes from initial damage caused by the pathogens, the positive feedback makes it difficult to see how such a response would end, other than in the death of the host organism.

Switching things around: if the host organism is able to do a similar trick, altering phenotypic traits to a wide variety of alternatives at no cost, then pathogens could not target them — any trait that could be exploited would quickly be altered to a non-exploited alternative.

We argue that ‘lock-in’ is vital to the kind of antagonistic co-evolutionary engagement that immune systems are part of. Without constraints, implicit in the substrate, restricting choice and imposing a non-zero cost of switching, selection pressure for antagonistic co-evolutionary traits is null. Adaptation is a favourable adjustment to some aspect of the world, but when the world is transient (because there is no cost to change) there is no gradient to climb: adaptations that impinge on other organisms cannot exist other than fleetingly. Such a world does not support predation or parasitism as a way of life, nor require, or provide selection pressure for, a reciprocal immune system.

This question may hold no great gains for the biologists, whose physico-chemical substrate of interest is known to support immunity, but this is where we contribute to the AIS community. If there are a class of substrates (problem domains) that fundamentally do not support the kind of antagonistic co-evolutionary engagement required for immune system evolution, then applying immune-inspired solutions to them may be misguided. A heuristic that determines whether a

---

2 An interesting thought-experiment. What would immune strategies look like if pathogens had no detectable appearance whatsoever? Perhaps the host could only detect damage in itself (e.g., heat shock proteins etc.) and change to a defensive state (e.g. inflammation) unfavourable to the pathogens.
problem domain is in the class to which immune-inspired algorithms might be usefully applied allows the correct tools to be selected for the job.

To make progress with this question, we will borrow from and extend a class of models known as \textit{NK} landscapes. These were conceived by Stuart Kauffman to investigate general properties of adaptive walks on rugged landscapes (Kauffman and Levin, 1987; Kauffman, 1993). Adaptive evolution, to a large extent, is a complex combinatorial optimisation process, and these models were designed to characterise the properties of the search spaces that adaptive evolution works upon. The size of the space, or fitness landscape, is the phase-space determined by \( N \) binary variables. A dependency network between the \( N \) variables determines how rugged the landscape is, how many conflicting constraints between variables have to be satisfied. The average number of dependencies per variable (or level of epistasis) is set by the parameter \( K \).

Two important assumptions are made here, concerning the \( NK \) model which apply to all three research questions. First, we assume that a binary phase-space (as used by the \( NK \) model) is sufficiently rich to represent problem-spaces ‘in general’ and that we can make general claims about our results. Second, we assume that a static fitness landscape model can represent what is, in reality, a much more dynamic optimisation problem. Both these issues can be addressed in part by an appeal to the literature – the \( NK \) model has been widely used to gain insight into complex optimisation problems in many domains, and its minimal representation contributes to that broad applicability. To the second issue, that we hope to make claims about dynamic optimisation problems based on results from a static optimisation model. A dynamic fitness landscape model couples a landscape for each population (e.g., hosts and pathogens) which deform one another as one population impinges on the fitness of the other (Bak et al., 1992; Hordijk and Kauffman, 2005). However, the questions we are asking here are more about evolvability and how deeply entrenched traits become so over evolutionary time. Intuitively, if we ‘fast forward’ a dynamic, shifting fitness landscape so it becomes a blur, only the most enduring features stand out and appear solid — these enduring features represent the optimisation problem we are interested in, that posed by the substrate, rather than the temporary environment. There is great scope to build upon this work later and incorporate a coupled landscape framework, such as the \( NKC \) model, but more of that in Chapter 5.

1.4.2 RQ2: Do more ‘biologically-inspired’ epistatic graphs exhibit greater ‘lock-in’?

By ‘lock-in’ we mean that some variables in the problem space have good values that are discovered early in evolution. These become fixed because their contribution is particularly significant. In biology, examples might include traits relating to energy gathering and storage, membrane building, or reproduction. A basic core of highly conserved traits representing the few
available solutions to these fundamental problems of life provides the lowest layer of building blocks, upon which exploration of other phenotypic dimensions may proceed. Bronowski (1970) termed this building up of layers ‘stratified stability’, that the lowest levels of complexity in biology must be relatively conserved, to serve as a foundation for higher levels to build upon. Dependencies, in the main, go down. Over evolutionary time, if these foundations are susceptible to changes in the environment, the entire phylogenetic edifice built on top of them will come tumbling down — great extinctions have happened many times in Earth’s history.

From the literature on biological dependency networks, which we will discuss in Chapter 2 and Chapter 3, we note the importance and prevalence of scale-free graph topologies. For example, the protein-protein interaction network in yeast (Fernández, 2007).

Our hypothesis is that the importance of the structure of the dependency network used in the NK model, usually having a simple ring or random topology, has been overlooked in the literature. We believe that constructing the dependency network so as to have a topology inspired by features of natural networks (arranging the K edges in a scale-free distribution) will cause increased lock-in of particular alleles in the resulting fitness landscape. Such locked-in alleles provide a stable foundation for further evolutionary exploration as Bronowski (1970) suggested, but also act as enduring signatures that are hard to substitute, allowing the organism to be recognised by a slower evolving immune system.

1.4.3 RQ3: Can a ‘biologically-inspired’ epistatic graph improve evolvability?

This question may sound rhetorical at first. If we observe scale-free epistatic networks in real biology, then we might expect that there is some adaptive advantage to them. If uniform or random network topologies were more robust, or had some other benefit, then surely evolution would have adopted them instead? Therein lies the question — if our assumption that the NK model is a suitable proxy to pose questions about evolvability and lock-in is correct, then we should expect that there are benefits to adopting a scale-free topology. Perhaps lock-in itself brings benefits (by reducing the search space at each step?) or perhaps there are other benefits to arranging the internal constraints within a system in this way, and the appearance of lock-in is manifestation of that. In either case, the existing means of assessing fitness landscapes, considering number and heights of local optima, and the sizes of their basins of attraction will give us a clear answer to this question.

1.5 Thesis structure

This chapter has introduced the general perspective we adopt on the evolution of immunity, and the specific research questions driving the thesis. Following this introduction, we review the
relevant literature on real and artificial immune systems, as well as touching on the constraints imposed by evolutionary biology in Chapter 2.

In Chapter 3 we restate the research questions in light of the literature review, and introduce and justify the choice of Kauffman’s \( NK \) fitness landscape model as a platform for the subsequent series of experiments. We further develop the hypotheses that structured epistasis will enhance evolutionary lock-in which has significance for immunological recognition, and that the apparent cost of lock-in will be accompanied by other measurable evolutionary improvement.

Chapter 4 describes the experiments using our development of Kauffman’s \( NK \) model. We report the method for creating scale-free epistatic networks (inspired by natural networks) and the problems in tuning such networks to arbitrary degrees of assortativity. We report a number of results from adaptive walks on the ensuing fitness landscapes, comparing our novel results with those from the literature where adjacent or random epistasis was used. Our results support the hypotheses — fitness landscapes from scale-free epistatic topologies have higher peaks, larger basins of attraction and global structures persist at higher levels of epistasis than previously known. In short, it is easier to reliably find good solutions when using evolutionary search on these new fitness landscapes — the ‘complexity catastrophe’ is mitigated.

Support for the second hypothesis, that such evolutionary improvements will be linked to an increased level of allelic ‘lock-in’ is less clear. A number of techniques are used to assess the entropy of individual alleles in the landscapes, but these results did not match with other landscape measurements. We conclude that, given the presence of global structures in these landscapes, it is not individual alleles that become ‘locked-in’, but co-adapted groups of alleles. Nonetheless, these groups represent entrenched features of the landscape, which provide a usefully conserved signal for immune recognition.

Concluding the thesis, Chapter 5 reviews the contributions made, critiques shortcomings of the approach, and outlines possible directions for future work.

1.6 Publications

The work in Chapter 4 follows on from our paper in the proceedings of Alife XI (Hebbron et al., 2008).

Other publications, on agent-based immunological modelling, that helped to develop the work found here appeared in the proceedings of ICARIS 2009 (Hebbron and Noble, 2009), and in the proceedings of ECAL 2009 (Hebbron et al., 2009).
Throughout most of human history the causes of disease were a mystery – a curse that could strike without reason and often did. Many schemes were invented to explain away, to tame this frightening phenomena, blaming various supernatural entities or sometimes more corporal sources such as the four humours: black bile, yellow bile, phlegm and blood. The true causes of disease were inconceivable before the discovery of microbiological agents. They existed in a world that was invisible both practically and conceptually.

Although the causes were a mystery, it was noted that some diseases, if survived, could leave the sufferer immune to future re-infection. The first commonly cited reference to this is by Thucydides, writing about the plague of Athens, in 430BC, during the Peloponnesian War.

Yet it was with those who had recovered from the disease that the sick and the dying found most compassion. These knew what it was from experience, and had now no fear for themselves; for the same man was never attacked twice–never at least fatally. And such persons not only received the congratulations of others, but themselves also, in the elation of the moment, half entertained the vain hope that they were for the future safe from any disease whatsoever.

In this background chapter, we review the history of ideas that has lead to our present-day understanding of immunology, briefly summarise the current textbook account of the human
immune system, and then situate this knowledge in an evolutionary framework. This perspective, on the present immune system as the product of a particular evolutionary history, brings us to the connection with evolutionary computing and cross-disciplinary work between computer science and immunology: the field of Artificial Immune Systems.

The models and contributions we present later in this thesis are made in this field — for researchers seeking to do computer science, inspired by the biological immune system. We want to understand the fundamental rules that underpin real immune systems, and the biological agents they control. Improving this understanding of the underlying real system domain can only help us when building computational immune-inspired models. We seek the deep and fundamental rules that govern the useful computational properties we observe embodied in immune systems. We want to extract the functions that transcend the biological substrate, for engineering, and separate the implementation details specific to the biological substrate. We want our pound of flesh, but not an ounce of blood.

Two diagrams (Figure 2.1 and Figure 2.2), developed from the literature, help us to orient the work to come within the Immunology / Engineering modelling space. The work presented in this thesis is a ‘methodological glue’ that helps to isolate a properties of systems to which immune inspired algorithms are applicable. Wolpert and Macready (1997) defined the “no free lunch” theorem. This states that, considered over all possible problem spaces, there is no general purpose optimisation strategy. Taking Artificial Immune System algorithms as a particular class of optimisation (or search) strategies, the community would benefit from knowing what subset of problem spaces these strategies do have broad application over. This thesis attempts to isolate one such signature that allows us to identify a class of problem spaces where AIS-type algorithms do have an edge.
Figure 2.1: An outline conceptual framework for a bio-inspired computational domain. Stepney et al. (2005) produced this outline of a conceptual pipeline for deriving bio-inspired algorithms from biological systems. We add the overlapping ‘necessary’ and ‘contingent’ elements within the biological system to emphasise that, when we sample their behaviour, it is difficult to separate the computational processes (of interest) from the contingent properties of the particular biological substrate that these processes happened to evolve in.

Figure 2.2: Space of models that capture a desired property of interest. Hart et al. (2013) capture the role of the Artificial Immune System practitioner as the arrows that move from immunological models of varying levels of complexity to the bottom right, to the minimal rational representation of a property of interest.
2.1 Immunology: a history of ideas

The science of immunology began, as with all sciences, with the observation of regularities in the nature of disease. Serendipitous coincidence allowed some early progress to be made against the infectious disease smallpox. However, it took the popularisation of the microscope in the 17th century and a long gestation period for microbiology to develop as a science. Accelerated by the pioneering work of Pasteur and Koch, microbiology then provided immunology with a conceptual and empirical foundation. Once researchers were able to identify and manipulate the agents of disease, the field of immunology developed rapidly, intertwined with the field of microbiology and later molecular genetics.

2.1.1 Smallpox and inoculation

The phenomenon of immunity to diseases was long recognised before the means and causes of transmission and infection were revealed. Plagues that swept through a population and killed many in quick succession were particularly feared, and often had a powerful impact on the course of history. We have already quoted Thucydides, describing the plague of Athens in 430 BC which killed a third of the population and was a deciding factor in the outcome of the Peloponnesian war between Sparta and Athens.

The disease in question is now suspected to have been smallpox, a highly infectious disease caused by the Variola virus. There are two variants of this virus, \( V. minor \) (\( \approx 1 - 3\% \) mortality) and \( V. major \) (\( \approx 30 - 50\% \) mortality). Normally spread by aerosol dispersion, coughs and sneezes, infection is also possible through person-to-person contact but is usually less deadly in this form as it does not reach the lungs. Symptoms include fever and a characteristic rash, becoming blisters which often left survivors of \( V. major \) severely scarred for life (Hopkins, 2002).

Features (antigens) common to both of the two variants allowed those infected with the milder \( V. minor \) to develop immunity to the more dangerous \( V. major \). The benefits of deliberate exposure to the milder disease to confer immunity against the danger of future infection were known in several ancient civilisations; in China as early as the 10\(^{th}\) century, and in India\(^1\) as early as 1000 BC (Wujastyk, 2001).

Inoculation (in general the deliberate introduction of one organism into the body of another in order to grow, more specifically called variolation when infecting a patient with \( V. minor \)) was performed in a number of ways: the powdered dried scabs of a \( V. minor \) sufferer could be blown up the nose, or the pus from a \( V. minor \) sufferer placed into a scratch on the arms or legs. By

\(^{1}\)In India the practice of gradually building tolerance to toxins in order to confer immunity was known and put to use in the form of beautiful ‘poison’ girls (vishkanya). These alluring assassins would then either administer a ‘kiss of death’, or share a poisoned meal with one’s enemy and survive (Singh Utpal et al., 2006; Sinha and Bhattacharya, 2006) — anticipating Ehrlich’s much later work inducing tolerance in mice to the poison ricin.
the 18th century, the practice had reached Turkey, and it was from here that Lady Mary Wortley Montagu, wife of the British ambassador to Constantinople, imported variolation to western Europe where, at the time, smallpox was responsible for ≈ 400,000 deaths a year (Dinc and Ulman, 2007). In Britain, the mortality was at times one sixth of the birth rate. Survivors were left scarred and disfigured: a third of all blindness at the time is attributed to smallpox scarring (Behbehani, 1983). With the support of the royal families and aristocracy, variolation slowly became established practice: the risk of the procedure (≈ 2% mortality) being far lower than that of contracting the more virulent smallpox.

In the 1760s and 70s a number of experiments were conducted, using pus from cowpox infections as an even safer substitute to V. minor inoculation. The best known of these experimentalists is Edward Jenner, a country surgeon who was teased by his friends that he should look for a milk maid as a wife, as they always had a clear complexion — never marked by pox scars. The virus responsible for cowpox, Vaccinia, shares a sufficient antigenic resemblance to V. minor and V. major that inoculation using cowpox does confer immunity against the whole Orthopoxvirus family of viruses. Inoculation using cowpox quickly replaced variolation as a far safer method of conferring immunity (Hopkins, 2002).

The technique of inoculation, for smallpox at least, was now reliable and safe, and the crucial aspects of artificial immunisation were revealed: inducing a mild version of a disease could confer immunity to later, more virulent infection, and immunity, once acquired, was preserved throughout the patient’s lifetime.

Serendipity played a large part in this early discovery: the widespread and devastating virulence of V. major coupled with the co-existence of less virulent members of the Orthopoxvirus family, and the established practice of variolation made an intriguing pattern for the prepared mind. However, it would be almost another century before germ theory provided a theoretical framework within which to explain this practical discovery, and thereby abstract the technique for use against other diseases.

The reader is directed to Riedel (2005) for greater coverage of Jenner’s contribution, and to other sources (Barquet and Domingo, 1997; Behbehani, 1983; Hopkins, 2002) for greater depth on the smallpox virus, the first to be globally eradicated (in 1979) by a concerted world-wide immunisation programme organised by the World Health Organisation.

2.1.2 Germ theory: the microbiological revolution

The fields of immunology and microbiology are inevitably intertwined — the mechanisms of immunity to disease can only begin to be conceived of once the disease-causing agents are identified. A great deal of our knowledge of microbiological life is driven by our desire to conquer
the spectre of bacterial infection, even though only a minute fraction of the microbiological world has any interaction at all with humans or our livestock, and it is a smaller fraction still that can be cultured ex vivo and studied in a laboratory (Stewart, 2012).

To begin to make any ground in the study of microbiology it was necessary to overthrow existing dogma: that life is spontaneously generated; that there is a ‘vital’ element to life, beyond mere physics or chemistry; the theories of four humours and of ‘miasma’ or bad air in health and disease. In short, to make medicine and microbiology empirical sciences, based on repeatable observations and not conjecture. Such observations were impossible until the invention of the microscope, late in the 16th century.

*Micrographia* was an early science best-seller: an illustrated account of microscopic structures, published in 1665 by Englishman Robert Hooke (Hooke and Jo. Martyn and Ja. Allestry., 1665). In it, Hooke coined the term ‘cell’ to describe a structure he observed, in cork, that reminded him of a monk’s accommodation. Most significantly for this account, it was also the first published depiction of a microorganism — the microfungus *Mucor*, a fact often overlooked when crediting Dutchman Antoni van Leeuwenhoek as the ‘first of the microbe hunters’ (Gest, 2004). Leeuwenhoek, a Dutch draper who created his own lenses for the inspection of cloth, was fascinated by the natural world exposed by these home-made, but powerful, microscopes. His letters to the Royal Society of London reported his many observations: moving single-celled creatures he named ‘animalcules’ (van Leeuwenhoek, 1702), red blood cells, sperm cells from animals and humans, and bacteria from between his own teeth. Importantly, he documented that maggots hatch from eggs, and did not appear by ‘spontaneous generation’, the prevailing belief at the time².

Despite these observations, the theory of spontaneous generation persevered – albeit now of cells or eggs rather than whole organisms. The related belief in ‘vitalism’, an elemental component of life beyond physics or chemistry, still clung on for another two centuries. Whilst subject to a number of challenges in this period,³ the definitive evidence against the theory did not come until 1859, when the French Academy of Sciences sponsored a competition to design an experiment to decide the issue.

Three years earlier, in 1856, a young Louis Pasteur, a chemist, had been asked to apply his scientific mind to a problem with industrial scale alcoholic fermentation. Thought to be a purely

---

²Unrelated to the story of immunology, but still fascinating, is that in 1674 Leeuwenhoek recorded having observed unexplained rapid random motions of non-living particles — a phenomenon finally documented and named, eponymously, by Scottish surgeon Robert Brown a century and a half later in 1827.

³Francesco Redi in 1668, compared meat in a jar, exposed to flies and subsequently found infested with maggots, with a control, meat in a jar successfully protected by gauze. John Needham, in 1745, experimented with boiled chicken broth in flasks, one open, one sealed – both were found tainted by microbes. Italian priest Lazzaro Spallanzani noted that Needham’s sealed jar had opportunity for contamination between the broth being boiled and sealed, so repeated the experiment with boiled, sealed flasks in vacuum, and no growth was observed. Critics claimed this was due to a lack of air, an argument countered by German zoologist Theodor Schwann in 1837, when he noted that fermentation did not occur in ‘sterile air’ (pure oxygen).
chemical process, it was nonetheless unpredictable. Pasteur was able to show that fermentation was a biological process, carried out by a living microbe, yeast, and that it was the presence of another microbe, a bacterium, that caused the frequent contamination. This “germ theory of fermentation” provided the insight necessary to design an experiment to win the 1859 competition.

Pasteur’s entry built upon previous experiments whereby nutrient broth was boiled in a flask, to ensure a sterile starting point, which was then sealed or exposed to air. His specially designed swan-necked flask allowed air to enter the sterilised body, and make contact with the nutrient broth, whilst trapping any airborne microbes in the neck. Sure enough, no bacterial growth was detected in the broth, but on tipping the flask so as to expose the broth to the trapped microbes in the neck, growth ensued. This experiment sounded the death knell for spontaneous generation, and promoted the idea that microbial life was ubiquitous, even borne on particles of dust in the air.\footnote{This confirmation, that all life comes from life, that there is no spontaneous generation, opens the door to important questions in the theory of life. Can we follow this chain of life-from-life backwards and ask questions about how it became so complex, and how it got started?}

The emperor Napoleon III (nephew of Napoleon Bonaparte I) had a problem with naval discipline — the rations of wine supplied to his sailors were often found to be spoiled and undrinkable. Pasteur, drawing upon his previous experience with microbes was able to identify the bacteria that caused the spoilage. Boiling, known to sterilise the wine, also ruined the taste. Pasteur was able to devise a method whereby he would heat the wine to a temperature sufficient to kill the microbes responsible for spoilage, without damaging the compounds responsible for flavour. He had invented the process of Pasteurisation — a cure for the ‘disease’ of wine. The following year, 1865, a disease in silkworms was crippling the French silk industry, and again, Pasteur was able to identify the infectious agent, detect infected silkworm eggs, and eliminate the spread of the disease (Debré, 1998).

If microbes could be the cause of disease in wine and worms, and sanitary conditions were the primary means of prevention, then why not apply this thinking to human disease too? This idea was enthusiastically embraced by the British surgeon Joseph Lister who sought out chemical methods (it not being possible to Pasteurise patients) to kill the microbes that infected wounds. Carbolic acid, known to prevent rotting in wood, proved to be a powerful antiseptic and “exercise a peculiarly destructive influence upon low forms of life” (Lister, 1867).

However, Lister’s insight was, sadly, too late to vindicate Hungarian physician Ignaz Semmelweis before his nervous breakdown and death, also in 1865. Semmelweis had, two decades earlier, noted that mothers were far more likely to die in childbirth from ‘childbed fever’ if they were attended on the teaching ward (where doctors had also been working on cadavers). Fatalities were far rarer on the midwives’ wards, or when giving birth at home, or even in the streets. Further evidence for the causal link came when in 1847 his friend Jakob Kolletschka died with ‘childbed
fever’ symptoms after being cut with a scalpel during an autopsy. Semmelweis instituted a routine of hand-washing with chlorinated lime which massively reduced mortality in childbirth. He published and championed his work, but it was ill-received, despite the evidence, as it implicated the doctors themselves in the transmission of disease.

Another significant and well-known event, again prior to the insight of germ theory, relates to a virulent outbreak of cholera in London in the summer of 1854. John Snow, a physician, traced the progress of the outbreak back to its source — a public water pump on Broad Street. His statistical methodology in retracing the transmission through the population, deducing that the causative agent was water-bourne, not due to ‘miasma’, established a means of scientific thinking about disease in a population, rather than in an individual, and he is regarded as the father of epidemiology (Snow, 1855; Cameron and Jones, 1983).

Taken together, these cases of disease in wine and in silkworms, where the causative microbes could be observed under the microscope, combined with the epidemiological evidence from Semmelweis, Lister and Snow, established a general germ theory: microbes could be the cause of disease in man, and explain its transmission within a population; by contact, by air or by water.

From the general to the particular, the next breakthrough was to identify the specific microbes responsible for specific diseases — scientific evidence for the germ theory. Robert Koch, a German physician, and a student of Friedrich Henle, who helped promote the germ theory, was prompted by Pasteur’s work on fermentation to investigate the microbe responsible for anthrax5. Living in the country, he had ample access to anthrax infected sheep and found the rod-like bacteria (*Bacillus anthracis*) in the blood of infected sheep, but never in the blood of healthy individuals. He also observed that sheep could be infected both by other infected sheep and by spores found in soil — the bacterium could go into a toughened state of suspended animation to survive a period of environmental stress outside of a host organism. To establish empirically the causal link between a particular bacterial strain and a given disease, Koch took infected blood and inoculated mice, who then died from the disease, and contained identifiable *Bacillus anthracis* which could, in turn, be cultured and used to transmit the disease to further host organisms6. The experimental procedure was generalised as Koch’s four postulates (Koch, 1880) establishing the causal relationship between a specific microbe and a specific disease:

1. The microbe is always present in each organism infected with the disease (but should not be present in healthy organisms).

---

5 Earlier in the 1800s, whilst the controversy over spontaneous generation remained undecided, the idea that anthrax was caused by microbes had already been proposed by a number of other scientists.

6 Koch and his laboratory assistants invented a number of enduring methods in microbiology during this period: the dye-staining of bacteria, making them easier to isolate and photograph; growing cultures of bacteria, first on slices of boiled potato, then in agar jelly; and the shallow dishes and covers, created by Koch’s assistant Julius Petri in which the cultures were grown.
2. The microbe must be isolated from a diseased organism and grown independently (in pure culture).

3. The cultured microbe should cause disease when introduced to a healthy organism.

4. The microbe isolated from the now-diseased experimental organism should be identified as identical to that isolated in the original source of the disease.

It was known at the time that this methodology contained flaws: for instance, Koch knew that there were asymptomatic carriers of cholera and of diphtheria, invalidating the universal claim of his first postulate, and viral causes cannot be cultured \textit{ex vivo}, as they require a host cell for reproduction. However, despite these caveats, the postulates provided a scientific framework for investigating many recognised diseases, and isolating their particular microbial cause. The fact that they are still referenced during disease outbreaks today, e.g., (Fouchier et al., 2003), is tantamount to their enduring value.

After anthrax, Koch turned his attention to identifying the microbes underlying human diseases, first septicaemia, and tuberculosis, the most significant public health problems of the day.

### 2.1.3 Immunisation

We return to Pasteur in Paris, now investigating chicken cholera using the \textit{ex vivo} culture techniques developed by Koch. Leaving for holiday one summer, Pasteur instructed an assistant to inoculate some chickens, which he failed to do, leaving the bacterial culture (\textit{Pasteurella multocida}) on a bench in the summer heat whilst leaving for holiday himself. After a month the now spoiled culture was used to inoculate chickens anyway, but they did not die as expected, and could not now be sickened, let alone killed by further injections of fresh cholera culture that still proved fatal to chickens who had not received the spoiled culture. Pasteur came to the conclusion that the spoiled, heat-damaged bacterial culture had conferred immunity to the chickens. He made the connection to the work of Jenner and smallpox, where a naturally occurring weaker virus conferred cross-immunity to the virulent smallpox causing \textit{V. major}.

Generalising from these observations raised the possibility of inoculation against any infectious disease: rather than needing nature to obligingly supply a weaker relative of the disease-causing agent, one could artificially weaken the original agent in its place, as had happened accidentally with the chicken cholera culture. Pasteur had glimpsed the biological basis for the practice of vaccination popularised by Jenner a century before, and named these artificially attenuated disease-causing cultures \textit{vaccines} in his honour.

No time was lost by the ‘microbe hunters’ in exploiting this process, spurred on by nationalistic and professional rivalry between Pasteur and Koch in the 1880s (de Kruif, 2002). Pasteur
produced a vaccine for anthrax in 1881 (*Bacillus anthracis* having been previously discovered by Koch). Koch, in 1882, isolated the *Mycobacterium tuberculosis* bacterium responsible for tuberculosis, the first specifically human disease to be targeted by the new methodology, but did not manage to create a vaccine.\(^7\) Pasteur set his sights on rabies. After animal trials, on 6\(^{th}\) July 1885 he reluctantly administered the vaccine to the first human subject: 9-year-old Joseph Meister who had been attacked by a rabid dog two days previously. The boy survived and was publicly held as further evidence for the microbial cause of disease in humans, and the efficacy of the science of vaccination.

### 2.1.4 Cellular and humoural immunology

As the new field of microbiology grew, attention turned to the finer details of interaction between the identified causal microbes and the cells of the host; how did bacteria damage host cells and tissues, and what was it that had changed in immune hosts that prevented this? The division between the two camps in Paris and Berlin continued, separated by their central focus; that the mechanisms of immunity were primarily cellular (Pasteur et al., in Paris), or primarily humoural — from factors in the blood (Koch et al., in Berlin).

Cellular immunology, the focus in Pasteur’s lab, was driven by the discoveries of Ilya Metchnikoff, a biologist, working on embryonic development in cuttlefish and crustaceans. Inspired by the engulfing and digestive processes of simpler organisms, Metchnikoff realised that this was not just a feeding response, but a defensive one too — unicellular amoebae cleared foreign materials by engulfing and neutralising them. Using starfish larvae as his (usefully transparent) animal model, he introduced a rose thorn and observed that it caused first inflammation and then a rush of white blood cells that did indeed attack and consume the foreign material. Metchnikoff also suggested that inflammation was not just for anti-microbial use — it could also be used (along with phagocytes) to cleanse and maintain the body: remove cancerous cells, inter-cellular debris, remove senile or effete elements, stimulate growth, and repair tissue through scarring (Gordon, 2008; Silverstein, 2011).

Humoural immunity, meanwhile, was founded by the discovery of bacterial toxins and antibacterial substances in blood. Friedrich Loeffler, in 1884, used Koch’s postulates to prove the suspected connection between diphtheria and the *Corynebacterium diphtheriae* bacterium, and also found that it excretes a soluble exotoxin. Alexandre Yersin\(^8\) and Emile Roux, in Paris,

---

\(^7\)The widely publicised discovery of *Mycobacterium tuberculosis* prompted great hope from the public for a cure, which did not follow. In 1890 Koch was disgraced by his attempts to profit from ‘tuberculin’, a dangerous and poorly tested vaccine containing dead *Mycobacterium tuberculosis* that had little therapeutic value. Nonetheless, Koch was awarded the 1905 Nobel prize in medicine for his work on tuberculosis.

\(^8\)Yersin was also the eponymous discoverer of the bubonic plague bacteria *Yersinia pestis* and its transmission via rats.
showed the exotoxin, not the bacterium, to be the causative agent by filtering to remove the bacteria: the filtrate still had the capacity to cause the disease.

Having isolated bacterial toxins in filtered blood serum, the crucial piece of the puzzle was to isolate the factors in blood that conferred immunity, and thus made immunisation possible. In other words, what were the active substances in blood that attacked or neutralised bacteria and toxins? Pasteur had used attenuated bacteria cultures successfully to vaccinate against chicken cholera and rabies, but the mechanism was still unknown.

Koch’s student, Richard Pfeiffer, had discovered a chemical, ‘endotoxin’, exhibited by bacteria that triggered an inflammatory response in the host, fever, and shock: the physiologically damaging symptoms of many diseases. These experiments revealed the existence of some humoral counterpart to endotoxin, capable of initiating a powerful, even dangerous, response from the body to a broad class of microbiological agents, based on a common attribute.

As the active substance causing disease could be transferred between hosts in serum, so too could the active agents of protection. George Nuttall had, in 1888, reported that blood had a broadly bactericidal action, but Emil von Behring and Shibasaburo Kitasato in 1890 showed that specific defences could be transferred in serum from an infected host and induce passive immunity in a second, previously unexposed, host. This process, of taking (what were later discovered to be) antibodies developed in one exposed host, and then using them in another host lead to the development of anti-venom and anti-toxins, initially against tetanus and diphtheria.

Bordet, in 1896, demonstrated that there are two factors working against bacteria in blood: a broad spectrum factor, inactivated by heating, and another, specific to particular bacteria, which survived heating. Separating these factors, he noted that the highly specific response did not kill bacteria in vitro on its own, it somehow enabled an enhanced response against its specific target, but that the killing action was performed by the broad spectrum factor. Whilst unable to kill bacteria when acting alone, he also noted the phenomenon of ‘clumping, not killing’, whereby many bacteria, or toxins, targeted by the specific response would become agglutinated in the sample.

Paul Ehrlich developed these ideas into an early version of the form we now know, naming antibodies, the specific factors which were somehow attuned to specific antigens with a lock-and-key mechanism, and complement, the heat-labile factor in blood that did the killing of bacteria.

---

9 Blood with all the red and white blood cells and clotting proteins centrifuged out, but still containing small exogenous and indigenous components: microbes, toxins, hormones, antibodies, etc.

10 We now know that the class of large molecules he had identified were lipopolysaccharides (LPS) an important and widely conserved constituent part of the outer cell wall in Gram-negative bacteria membranes, and consequently a widely recognised pathogen-associated molecular pattern (PAMP) in immune systems, from plants to vertebrates.

11 Anti-venom, antidote, or anti-toxin are used when a patient doesn’t have time to generate their own immune response. Donated antibodies bind to antigen on the pathogen or toxin and allow the immune system to ‘see’ the unfamiliar antibody and mount an immediate response to it, rather than waiting for their own adaptive immune system to discover the antibodies for itself.
Ehrlich’s ‘magic bullets’ — the diverse antibodies that were so finely tuned to specific antigens — caused him some concern. He realised the potential for ‘horror autotoxicus’ if antibodies were generated that aimed at patterns of the host organism itself. He reasoned that the body must have some means of preventing this occurrence. He also speculated, some three decades before the structure of antibodies was revealed, how the matching region, which he called ‘side chains’, might be generated by cells under attack, corresponding to patterns on the attacking entity, and released as free-floating antibodies in the blood to bind to more of the same.

The link between these specific sensors and the killing factor in blood, the complement, was uncovered by Almroth Wright in 1904, and named opsonisation; a tagging mechanism by which antibodies bind to their selected epitope (binding region of antigen) with their variable region (or side chain) and leave the fixed tail (Fc region) visible as a marker sticking out from the foreign body. Complement, a chain of chemical responses, can be initiated by this marker, attaching to the Fc tail and eventually punching a hole in the bacterial cell membrane, effectively killing it (via the small proteins of the membrane attack complex). Some cells are able to resist this process, and for some cells it would be better to clean them up more carefully rather than to burst them and thereby leak toxins into the body. Macrophages sport an Fc receptor for these cases, and, as Bernard Shaw put it, pathogens are “buttered, appetizingly” for phagocytosis (Silverstein, 2011).

This revelation, that both humoural and cellular immune components interact, and the joint award of the 1908 Nobel Prize in Physiology or Medicine jointly to Mechnikov and Ehrlich, rejoined the two factions, and set the scene for modern immunity. From this beginning, the twentieth century saw great leaps of development, both in the understanding of the cellular and molecular components of the immune system, whose foundations we have covered, and in the corresponding features of our pathogens, from bacteria to viruses, parasites and fungi. The discovery that many bacteria and fungi themselves produce broad-spectrum antimicrobials that can be harnessed as therapeutics heralded the age of antibiotics (perhaps only a momentary upper hand in the war between us and microbes, thanks to irresponsible overuse) and the age of genetics allowed the mechanisms within and relationships between our pathogens to be revealed for the first time, and even targeted systematically (Barabási et al., 2011; Goh et al., 2007).

Whilst immunological theory had, at the turn of the 20th century, reunited the humouralists and the cellularists, the vast majority of discoveries during the first half of the 20th century were in the mechanisms and applications of antibody generation and antibody / antigen binding. How did the body produce the enormous diversity required, how did it avoid ‘horror autotoxicus’, and how did recently identified viruses fit into the picture?

Viruses posed a problem for early microbiologists as they could not be seen but could be detected as infectious agents. They passed through filtration (showing them to be smaller than bacteria), but were not toxins as they reproduced — although only in cells, so they could not be cultured.
Martinus Beijerinck first reported the existence of a virus in 1898 (tobacco mosaic virus). Another notable event in the history of immunology was the discovery that a transmittable virus could alter cells and cause cancer, shown by Peyton Rous in 1911.

Karl Landsteiner observed, in 1900, that blood from two donors agglutinates when put together, and deduced that this was because differing antigens on red blood cells have corresponding antibodies in serum. This led to the identification of blood groups (determined by the red blood cell antigens present), and subsequently to successful blood transfusion (wherein the blood group must be a match, i.e., one must ensure that the antigens are the same between donor and recipient). Landsteiner went on to perform further experiments with the immune system, in 1917 demonstrating that it was possible to sensitize, i.e., to generate antibodies, in response to synthetic substances (haptens) as well as to bind to and effectively block specific antigens by the same method. This provided a new experimental tool for investigating the immune system and its interactions with toxins, pathogens and the self (Shreder, 2000).

Linus Pauling, in the 1940s, used Landsteiner’s methods to perform experiments on the structure of antibodies in order to better understand protein structure and the protein-protein interactions that he believed were crucial to biological processes. He confirmed the lock and key, or hand-in-glove model, i.e., that complementarity operated by molecular shape, rather than by chemical bonding. The implications of biological complementarity and weak, reversible bonding went far beyond the immunological case, to a much broader theory of biological specificity with application to very general enzymatic binding to substrates, as well as other specific cases such as olfactory sensors.

By the mid 20th century, it was clear that, despite the great progress in understanding the humoral antibody system, there were gaps in the theory. The presence and action of antibodies alone did not explain the delayed-type hypersensitivity to some antigens (first noticed by Koch when working on Tuberculosis in the 1800s), or the rejection of allograft transplants described by Medawar (1944). Conclusive evidence that adaptive immune responses also employed a cellular component, as well as humoral antibodies, came from experiments on guinea pigs, carried out by Karl Landsteiner and Merrill Chase in 1942. Lymphatic cells taken from a guinea pig infected with Mycobacterium tuberculosis were transferred to a second naive host guinea pig. When this second host was inoculated with the antigen, it elicited an immune response which did not occur in other naive, uninoculated hosts, and this reaction could not be elicited by the transfer of serum, so could not be due to antibody transfer. The lymphocytes responsible were later identified by Gowans et al. (1962).

As well as the reintroduction of a cellular component to acquired immunity, there remained the sense that antigens directed the formation of antibodies; that the antigen somehow acted as a template. This paradigm was overturned by Jerne in 1955 when he introduced his clonal natural selection theory — a diverse variety of globulins (antibodies), generated by some unknown
method, are selected by matching antigens whereupon they are transported to antibody producing cells that copy the selected antibody. The source of the diversity was unknown, and the means of proliferation once selection occurred was incorrect, but the Darwinian process that amplified the response to matching features was revealed (Jerne, 1955). Frank Macfarlane Burnet and David Talmage then placed the lymphocytes that produce the variable antibodies as the selected-for and proliferating components and put forward the idea that some selected and circulating lymphocytes might be long-lived enough to provide the observed property of immunological memory (Burnet, 1959; Talmage, 1959).

Burnet went on to develop the idea of a gradually learned immune response. As, demonstrably, some immunity was learned, and not implicit, perhaps all immunity was not innate but acquired over our lifetime. For this process to work, and recalling Ehrlich’s ‘horror autotoxicus’, the immune system must also learn to avoid damaging the self. Burnet thereby introduced the immunological concepts of self and non-self, and, along with Lederberg (1959), the related concept of central tolerance, by which the immune system comes to recognise the self and any lymphocytes that are in danger of becoming self-reactive are deleted from the repertoire (Park, 2006).

The genetic mechanism (VDJ recombination) responsible for the diversity of antibodies was discovered in the 1970s, earning Susumu Tonegawa a Nobel prize. Hozumi and Tonegawa (1976); Tonegawa (1983).

Jerne (1974) introduced another new paradigm, the immune network theory, that briefly states that not only do lymphocytes recognise antigen, but they also recognise one another. This network of interacting lymphocytes operates as a complex system, regulating specific nodes and supporting emergent features — the nebulous dynamic immune ‘self’.

Many more discoveries than we could hope to summarise were made as the 20th century progressed; we now move to a contemporary overview of the immune system, placing these cells and components in context.

### 2.2 Modern immunology

The development of the field of immunology leads us to the current orthodox view of the immune system in humans: the textbook account. We say orthodox, as the prevailing view is still very much of the self-non-self school of thinking, although more recent challenges to this perspective are discussed below. We will briefly introduce the tissues, cells and molecular components of the immune system, before considering their evolutionary history. A number of textbooks and review articles have been invaluable in writing this section, notably Murphy (2007); Kindt et al. (2006); Clark (2008); Playfair and Chain (2012) and Delves et al. (2006).
Chapter 2 Background

The human immune system is commonly regarded as consisting of two sub-systems, the innate immune system, and the adaptive, specific or acquired immune system. The innate system is older, its capacity for recognising and eliminating pathogens are encoded in the genome, and many features are highly conserved across plants, insects and other animals. The adaptive immune system is much newer, found only in higher vertebrates. With its capacity for learned immunity it provides the medically interesting possibility of immunisation: provoking the recognition, memory and elimination of pathogens not recognised by the innate system. With this in mind, it is not surprising that the adaptive system has been of primary interest historically. In the last two decades this bias has changed, and the role of the underlying innate system, and the ways in which the two arms of the immune system interact is increasingly of interest (Iwasaki and Medzhitov, 2010; Germain, 2004; Medzhitov, 1998; Flajnik and Du Pasquier, 2004).

2.2.1 Adaptive immune system

The adaptive arm of the immune system consists of lymphocytes, the white blood cells with their diverse and highly specific receptors, along with organs and processes to finely tune these receptors to cover the space of possible patterns (epitopes) presented by antigen.

Interactions between lymphocytes and signalling molecules corroborate highly-specific antigen detection and initiate pathways that alter the bodily environment, tag pathogens for elimination, or damp an immune reaction in progress. Elsewhere we have described the adaptive immune system as a ‘sensor upgrade’ for the more ancient innate-immune system and this is a useful analogy to use – a more complex organism that moves in more varied environments encounters more varied antigens. Assimilating the information to recognise all this diversity into germline sensors is costly, perhaps dangerous. The adaptive immune system solves this problem through dynamically adding to the receptor repertoire during the hosts lifetime, and passing the information to the existing immune infrastructure to deal with.

All lymphocytes, white blood cells, are derived from common progenitor cells, and differentiate into three classes: NK (natural killer) cells destroy host cells that are obviously infected or damaged; B-cells produce antibodies and perform extra-cellular pattern recognition; and T-cells are responsible for intra-cellular pattern recognition — particularly important for handling viruses, as these are only biologically active once they are inside a cell and out of view from the humoural components of the immune system.

To perform the broad spectrum, highly specific recognition of epitopes, the receptors of the adaptive immune system are constructed pseudo-randomly from a library of sub-gene fragments. This process, somatic recombination, is also called VDJ recombination, after the Variable, Diverse
and Joining elements. The combinatoric approach allows the generation of a diverse range of receptor locks\textsuperscript{12} to match an even larger space of keys (epitopes).

This ‘brute-force’ approach naturally generates some receptors that match self-epitopes – features of the host. Lymphocytes that have receptors that bind too strongly to self-epitopes during development undergo apoptosis, in this context called clonal deletion, part of the process of ‘central tolerance’: ensuring that self-reactive lymphocytes do not leave the primary lymphoid tissue where they are initially created.

As T-cells have to ‘see’ inside other host cells for signs of unusual activity, a mechanism for presenting samples of intracellular proteins is required. All cells regularly transport peptides to the surface, bound within an MHC molecule\textsuperscript{13}. For most cells this is MHC class I, but in specialist antigen-presenting-cells it is MHC class II, signalling that the presented peptide sequence originates from disassembled antigen. Central tolerance for T-cells (in the thymus) involves two phases: positive selection for MHC binding, and negative selection for MHC+self-peptide binding.

Having passed central tolerance, B- and T- lymphocytes pass into circulation, waiting to encounter their specific antigen.

When antigen binds to a B-cell-receptor with sufficient affinity, it is absorbed into the B-cell and digested into its constituent peptides (see Figure 2.3). These constituent parts of the antigen are then wrapped in an MHC class II ‘portal’ molecule and transported to the surface of the B-cell, presenting an MHC+peptide target for a corresponding T-cell-receptor. This allows the adaptive immune system to make the correlation between antigen surface features (a non-self epitope matched by a B-cell-receptor) and what it’s made of (the non-self peptide recognised by the T-cell-receptor).

A B-cell that has both signals, bound antigen, and co-stimulation from a T-cell, is activated\textsuperscript{14} and begins cell division to expand the number of cells specific to its matched antigen. The vast majority of the resulting clones become short-lived plasma B-cells that secrete huge numbers of antibodies to quickly clear the infection by binding to and tagging the antibody for elimination.

A minority of the activated B-cell clones migrate to a lymph node, to a germinal centre\textsuperscript{15} to become memory B-cells (Gatto and Brink, 2010). Here the activated B-cells continue to proliferate but also undergo somatic hypermutation – regions of the cloned B-cell-receptors

\textsuperscript{12}Paratopes, the Y-region in the immunoglobulin protein
\textsuperscript{13}Major Histocomptibility Complex family of molecules, or Human Leukocyte Antigen family in humans.
\textsuperscript{14}There are mechanisms for T-cell-independent-activation – these all involve some form of secondary activation from another receptor, for instance complement (Carroll, 2000) (an important link between the adaptive and innate systems) or from multiple co-stimulation from repeating antigenic patterns, e.g. the flagellum polymers that make up a bacteria’s tail (Vos et al., 2000).
\textsuperscript{15}There is some evidence that B-cell somatic hypermutation does also happen outside germinal centres (William, 2002)
exhibit a mutation rate far higher than the normal, leading to variation in affinity for the original antigen. Selectively, those clones that have higher affinity to the antigen are retained: a Darwinian process of exploring the feature-space around the original antigen. Selected memory B-cells live for many months, and divide without needing further activation from antigen. On a second exposure to the same antigen the immune response is rapid, and highly effective – the host has a memory of the previous infection, and has acquired immunity to subsequent encounters. The ability to induce this process of immune memory formation through deliberate exposure to antigen (plus an adjuvant to kick-start an immune reaction) forms the basis of vaccination.

2.2.2 Innate immune system

Prevention is better than cure: stopping any potentially dangerous microbiological agent from ever entering your body where it might do harm is the best defence. The front line of the innate immune system is the skin: a basic physical barrier to the entry of pathogens into the body. Necessary entry points to the body are coated in mucous membranes and antimicrobial secretions to deter pathogens, and tolerated colonies of benign bacteria help prevent more pathogenic strains from taking hold. Another line of innate defence is the internal environment of the host. Microorganisms are, for the most part, very sensitive to their environment, particularly when it comes to the conditions required for growth. Temperature and pH in the body, for those microbes that manage to enter, make it an unsuitable environment for many (but not all) species.

At the centre of innate immunity is the complement system, first discovered by Buchner and Bordet (so named because its action in eliminating bacteria complemented the detection provided by antibodies). This collection of proteins in blood plasma provide multiple chemical pathways from receptors to effectors, historically classified into three pathways:

**Classic pathway** triggered by antibodies that have encountered and bound to their corresponding pathogenic antigen\(^{16}\).

**Alternative pathway** triggered by certain Pathogen-Associated-Molecular-Patterns (PAMPs), but most importantly it is a ‘fail-dangerous’ mechanism – the absence of control molecules from healthy self cells trigger this pathway in the presence of altered-self or non-self.

**Lectin, or Mannose-Binding-Lectin pathway** triggered by certain sugars, specifically non-protein surface features of bacteria and fungi.

The resulting cascade, once triggered, is amplified massively and initiates the production of anaphylatoxin peptides to induce inflammation; performs opsonisation (tagging) of pathogens;

\(^{16}\)The classic complement pathway is an example of the evolutionarily later adaptive immune system ‘hooking in’ to more ancient innate immune responses.
attracts phagocytes to the area; and finally produces the membrane-attack-complex, a tube-like complex that punches a hole in the membrane of pathogenic cells, fatally leaking their contents (lysis).

As with much work on biological systems, further research reveals deeper complexity. Ehrn-thaller and Ignatius (2011), in a review of the complement system, concluded that the three canonical pathways are too simple a model of the complex interactions between the complement components, and other systems such as the coagulation cascade.

![Figure 2.3: Phagocytosis. A target binds to receptors on the surface of the phagocytic cell (a), which then deforms to engulf the target (b). The target is encapsulated in a phagosome (c), which then fuses with lysosomes (d), containing enzymes that digest the target (e). To remove waste the phagocytic cell can reverse this process: fusing the phagosome with the cell surface and expelling the contents (not shown). Targets may be innately recognised, highly conserved pathogen-associated-molecular-patterns, or tagged by antibodies for opsonisation.](image)

The cellular elements of the innate immune system fall into two groups – the phagocytic cells that ingest foreign material (Figure 2.3), particularly in response to opsonisation signals tagged to cell surfaces, and the granulocytes (neutrophils, eisinophils, basophils, and mast cells) which contain granules of pathogen-class-specific toxins, or reactive molecules to initiate the inflammatory response, dilating blood vessels and recruiting further immune reaction.

Recognition in the innate immune system is performed by germline encoded pattern recognition receptors (PRRs). As these receptors match highly conserved features of pathogens, it should perhaps be no great surprise that very close homologs to the human PRRs have been widely discovered in other organisms. The Toll-like receptor family is notable in this regard, having been identified in animals, insects, and plants (Medzhitov et al., 1997; Leulier and Lemaitre, 2008; Ronald and Beutler, 2010). Furthermore, there is evidence that an ancestor of the Toll-like family of receptors was already well entrenched half a billion years ago among the Bilateria – progenitor of both insects and mammals (Beutler and Poltorak, 2001).

However, pathogens are not the only cause of damage to the host organism. Trauma, causing tissue-damage and necrotic (unplanned) cell-death all release substances into the extra-cellular space that would normally only exist within a cell. These molecular patterns, dubbed Damage Associated Molecular Patters (DAMPs) also provide import signals to the innate immune system (Bianchi, 2007).
2.2.3 Theory

As knowledge of the components and interactions of the immune system has grown, new theoretical frameworks have developed to make sense of this complexity – to give us a systemic understanding.

In the past few decades, new theoretical models or perspectives have been proposed on several fronts, for instance, idiotypic networks (Jerne, 1974), and the cognitive immune theories of Cohen and others (Cohen, 1992; Hershberg and Efroni, 2001).

Explanatory problems with the self-non-self paradigm were increasingly apparent at the end of the 20th century. The established clonal selection and central tolerance theories of Burnet (1959) that, with some alterations, remained the paradigm for orthodox immunology were increasingly challenged by awkward experimental results that did not fit the theoretical framework:

1. Vaccinations require adjuvants (a secondary signal is needed to prime the system to bother recognising the attenuated or denatured virus or bacterial sample you are presenting to it.)

2. ‘Self’ changes over time (puberty, pregnancy, milk proteins) not all self-proteins can be selected against in the thymus. Also, tumours are not rejected.

3. We rely on symbiotic bacteria in the gut, in the mouth, on the skin – everywhere!

4. Auto-immune diseases exist (MS, arthritis, allergies, etc.)

Matzinger (1994) introduced and developed her novel “danger theory” about the most central function of the immune system (Gallucci et al., 1999; Matzinger, 2001, 2002, 2007), to contend with these challenges to the orthodox model, even after a number of attempts to patch it (Janeway, 1989, 1992). The development of self-non-self and derived models, the danger theory and the partitions on antigen space they operate on are illustrated in Figure 2.4.

Matzinger proposed that the immune system’s central job was not to react to self or non-self, but to react to evidence of damage or danger in the body. Thus, the presence of a pathogen would not be detected just by recognizing its shape but by recognizing what it had done. Alarm signals and the presence of chemical components normally found only within a cell constitute evidence that something unusual (i.e., dangerous) is happening, and it was Matzinger’s argument that these signals and not template-matching prompted defensive reactions on the part of lymphocytes. To return to the city metaphor with which we opened the thesis, this equivalent to suggesting that the city guards do not keep a mental list of the faces of wanted criminals, they just approach anything that sounds like a fight or a robbery and then dispense justice when they get there.
Figure 2.4: Partitioning the universe of antigens — adapted from (Matzinger, 2002). The self-nonself (SNS) models divide the space into self (a) and non-self (b). The infectious nonself (INS) models into non-infectious self (a) and infectious-nonself (f). The danger model changes the focus to danger-associated signals: altered self (c), non-self and dangerous toxins (d), and dangerous non-self (with conserved pathogen associated molecular patterns) (e). The theory also has the advantage of distinguishing non-dangerous non-self (f), even those exhibiting PAMPs, which include symbiotic organisms such as gut bacteria, a problem for previous theories.

2.3 Biosemiotics

Danger signals are not unique to our immune systems: Witzany (2006) describes plants signalling to each other using small molecules and proteins common across whole families and genera. When leaves are damaged, volatile compounds are released that either deter parasites, attract the natural predators of the parasites, or warn nearby plants allowing them to prepare a pre-emptive response. (This is an ancient strategy for communicating through the environment that bacteria also use, i.e., quorum sensing.) These suitably named ‘semiochemicals’ become a feature of the world worth tuning into — a source of adaptive information about events in the locality (Paré and Tumlinson, 1999).

Along related lines, the early ethologist von Uexküll (1992, original 1934) introduced the notion of the umwelt. This term refers to the subjective perceptual universe of an agent. Uexküll’s feedback-cycle model for biological meaning anticipated cybernetics (the study of control systems) by decades. The field of biosemiotics developed these ideas, supported by the semiotic logic of the
philosopher Charles Sanders Peirce (Sebeok, 2001). In Peirce’s view, logic is a formal doctrine of signs. For Peirce a sign relationship is more than just the sign and its meaning (e.g., red blood and injury). It is a triadic relationship between the sign vehicle (epitope), the thing (pathogen), and the receiver (immune system). It is also noteworthy that the receiver of the sign has to be prepared in some way to understand it.

The immune system, with its rich fabric of inter-cell, inter-species signalling, information, concealment and learning is of great interest to the biosemiotics field (Neuman, 2008). In a provocatively titled paper, Hoffmeyer (2011) suggests that all ‘biology is immature biosemiotics’ – that without employing semiotic terms it would be impossible to describe function at all. We will return to the question of function, as the ultimate explanation for a biological trait in Section 2.4.1.

For a ‘molecule to become a message’ (Pattee, 2012; Umerez, 2009) requires a degree of stability, of information currency, to hold for long enough for evolving agents to learn the correlation between the molecule and the adaptive value of the message it represents. Completely contingent signals are ephemeral: if they disadvantage the signaller the molecule can be altered and the signal lost. When, however, a distinct molecule is unavoidably produced and released, e.g., as a secondary metabolite of some critical and well-optimised cellular process, then this signal is much harder to conceal – it is functionally constrained. Such signals represent useful features for a detector – a useful detector needs a simple signature to reliably represent a more complex system. To understand how such signals become established, and even entrenched (Scheffer and Westley, 2007), we must turn to evolutionary biology.

2.4 Evolutionary biology

When Charles Darwin published his theory of evolution by natural selection in 1859 (Darwin, 1859) he expressed what must be the single unifying idea of biology. Centuries of observations and evidence from the natural world were synthesised into a single theory. The well-adapted organisms we see today come about through a history of populations producing many offspring, some with variations, these variations are in turn inherited, and that natural competition, over time, selects for variations that give an advantage in survival and reproduction.

It was, perhaps, an idea that came of its time – this is supported by the concurrent work of Alfred Russel Wallace, with whom Darwin famously shared publishing priority (Darwin and Wallace, 1858). Many ideas of the time can be seen as contributing to the theory – the competition from population explosions that concerned Malthus (1826), the ordering and classification of the natural world performed by Linnaeus (1758), existing ideas on evolution and adaptation from Lamarck (1809) and the evidence from fossils and ideas about continental drift for a deep
biological history and of long-term geological change – time enough for evolutionary adaptation to arrive at the “endless forms most beautiful and most wonderful” (Darwin, 1859).

The rediscovery of Mendel’s work on inheritance patterns of peas (Mendel, 1865), in the early 20th century provided the unit of inheritance missing from Darwin’s work, but caused a rift between Mendelians, who claimed that mutations were the source of variation and the starting point for new species; and the Biometricians, who claimed that mutations are irrelevant noise, quickly lost – natural selection operates on the continuous variation in the population.

The two fields were eventually reconciled through the works of statistician R. A. Fisher and geneticist J.B.S Haldane, who demonstrated that the continuous variation the Biometricians favoured could be generated by the combined action of mutations at many different genetic loci: natural selection could alter allele frequency within a population, resulting in altered traits over many generations and, ultimately, evolution (Bowler, 2009). Famously, this theoretical breakthrough was illustrated by the case of the rapid evolution of the peppered moth from a light coloured to a dark phenotype – better to maintain camouflage in the increasingly grimy environment of the industrial revolution.

Wright (1932) introduced his adaptive landscape metaphor to quickly explain his shifting balance theory in a less mathematically dense way – to explain how populations could move between alternative peaks on a rugged surface caused by interactions between loci. Wright restricted his landscapes to two-dimensions to demonstrate his point, but the metaphor has endured. One criticism is that our intuitions about movement between valleys and peaks in two or three dimensions are very different to the reality of movement in high-dimensional spaces (Kaplan, 2008).

The new synthesis in the early mid 20th century set evolutionary theory on a course to uncover the mechanisms of inheritance and adaptation, and ultimately to reveal the secrets of DNA, life as information, and to ask questions about the abiotic origins of life on this planet.

2.4.1 Evolutionary explanations

When we observe some biological trait in an extant organism, and begin to wonder how and why it has a certain shape, uses a particular material in its construction, or behaves in at specific way we are asking several different types of question about the trait. The ethologist Tinbergen (1963), developed a schema of four categories of enquiry, originally for evolved behaviours, but equally valid and relevant for other biological or ecological traits (Bateson and Laland, 2013):

**Mechanism**: how does the proximate system work? What are the component parts, and how do they interact?
Ontogeny: in an organism, how does development of the mechanism proceed from origin to maturity?

Phylogeny: what is the evolutionary history of the mechanism and its ancestors? What innovations appeared, what scaffolding disappeared? What is the particular path taken through the space of possible phenotypes by the ancestors of the modern organism?

Function: what is the selective advantage in being equipped with such a mechanism? What ecological problem does it solve? The adaptations in the phylogenetic route taken by a species are specific responses to selection pressures. A truly explanatory account of an evolved mechanism must address the deeper structure of the problem for which the specific mechanism is one possible solution.

Understanding the selective forces that help explain the current nature of a trait can be very difficult, but it is often the most engaging question (Nesse, 2013). As well as their engaging nature, questions of function, we argue, are the most relevant when trying to separate the computationally useful functionality of a biological system from the legacy of its evolutionary history.

Whilst evolution may appear to produce general purpose solutions, this is not the case; solutions in real biological systems are *nichiversal* (not universal). Organisms evolve to solve the problems their ancestors faced, but not to solve a general class of problem (Bullock, 2006). Solutions are tightly bound to the context in which they arose, and represent a compromise balancing multiple conflicting constraints on the organism. It is this history of specific challenges driving immunological inventions that we consider next.

### 2.4.2 Evolutionary history of immune systems

Although our aim is to develop functional questions, and explanations, which transcend the specific phylogenetic history of biological immunity, it is the only history we have to hint at that underlying structure. Following the approach of Maynard Smith and Szathmary (1998), we describe here a sequence of major transitions in the evolution of immune systems.

The first auto-catalytic replicators to control their local environment within a compartment, or primitive membrane, and to replicate as an individual represent the first agent that might have had sufficient agency, sufficient selfhood, to have need for a sense of self vs. non-self. (Hoffmeyer, 1998).

Prokaryotes, single celled bacteria and archea, have long dominated life on Earth. Even at the lower end of the complexity scale, they are not too simple to avoid being threatened by their own pathogens. Bacteriophages are viruses that replicate in bacteria, which naturally have evolved methods to resist infection (Weitz et al., 2005; Labrie et al., 2010). Phage therapy
is of practical interest for controlling bacteria in medical (Kutter et al., 2010) and industrial applications (Petty et al., 2007), particularly in the wake of widespread antibiotic resistance (Pirnay et al., 2011). Prokaryotes also take in mobile genetic elements, and have evolved a pattern matching and rejection mechanism to protect themselves against those that interfere with their normal operation: the CRISPR (clustered regularly interspaced short palindromic repeat) system (Brouns et al., 2008; van der Oost et al., 2009; Seed et al., 2013). A much more recent discovery than phages, there is also great interest in exploiting the CRISPR system, which has no equivalent in eukaryotes, in research, medicine and industry for its ability to ‘vaccinate’ bacteria to recognise specific patterns and act upon them.

Lane (2005) writes a compelling account of perhaps the singularly most unlikely event in the history of life on Earth – eukaryogenesis. A prokaryote acquired symbiotic mitochondria and escaped from the energetic canyon imposed by the ratio of surface area (energy generation) to volume (energy consumption) (Cavalier-Smith, 2002; Yutin et al., 2009). This event marked the crossing of an almost impossible valley in the fitness landscape – forever dividing the eukaryotes from the prokaryotes and making prokaryote cell surfaces inescapable markers that eukaryotes have escaped, and that many immune receptors would come to recognise.

The freedom to grow much larger and more energetic, and motile, (trends that would be recapitulated later by much larger organisms) enabled new ways of life and increased consumption of resources. The origins of predation and anti-predation responses are entwined with the advent of eukaryota (Cavalier-Smith, 2009; de Nooijer et al., 2009; Jurkevitch, 2007a,b).

As predatory interactions developed in unicellular life, defences were evolved to maintain the balance. At some point, the costs of defensive phenotypes were too costly to bear continuously. Coupled with continued predation this provided the evolutionary pressure for a switch from passive defences, to more active and inducible defence activated by altered pattern receptors that would previously recognise food (Poitrineau et al., 2004; Wiackowski and Staronska, 1999; Zangerl and Rutledge, 1996).

The advent of multicellular life brings its own unique challenges: recognising a distributed self is far more difficult in a colony (e.g., a colony of sponges) where your genetic destiny is shared by your kin, and a similar but genetically-more-distant neighbour next to you is so similar that you can’t tell the difference, the neighbour can deceive you into providing whatever services you provide to your colony, for them. Being able to identify very close but not identical kin can be as important as identifying pathogens. Another example here is tumour cells — they of course are ‘self’ cells, but have slipped their program — the requirement to detect defector cells is very strong, and the many layers of mechanisms evolved to yoke self-destruction and signalling to damage-detection and defection show this. Finally, on multicellularity, we see pressure to develop apoptosis, controlled and programmed cell death for the benefit of the organism (Grosberg, 1988; Dishaw and Litman, 2009; De Boer, 1995; Rinkevich, 2004; Heck et al., 2005).
As we discussed in Section 2.2.2, by the time, some 1 billion years ago, that the last ancestor of plants, insects and animals was alive, what we might recognise as an innate immune system was already established: pattern recognition receptors coupled to inducible defences and highly specific self identifying patterns within a multicellular host. Post Cambrian explosion we suddenly see the emergence of acquired immunity: the origins of the sophisticated sensor upgrade the immune systems of jawed vertebrates acquired approximately 500 million years ago (Matsunaga and Rahman, 1998; Laird et al., 2000). This enabled them to adapt to pathogens within a single lifetime, rather than over evolutionary time as with the innate immune systems that all living things have. (Schmid-Hempel, 2003; Edelman, 1994; Matsunaga and Rahman, 1998; Laird et al., 2000).

Why did the adaptive immune system evolve from the earlier innate system? The adaptive system, providing acquired immunity, is often held to be the pinnacle of immune defence mechanisms: a high-tech, high specificity ‘sensor upgrade’ for the comparatively broad spectrum sensors that suffice for lesser organisms. It is, supposedly, an evolutionary step taken exclusively by the complex jawed-vertebrates to cope with correspondingly complex threats (Travis, 2009). However, this may be a somewhat anthropocentric view: plenty of organisms survive for a nice long life without an adaptive immune system (Schmid-Hempel, 2003; Hedrick, 2009). Perhaps short-term, myopic selection for an adaptive system gave a short-term boost against pathogens, but was a one-way ticket: once an adaptive system exists, it may drive selection for greater pathogen evolution. In other words, the red queen effect (Van Valen, 1973) steps up a gear, but still doesn’t get anywhere new in terms of the overall fitness costs to the host (Hedrick, 2004). As discussed earlier, that the adaptive immune system arrived approximately 500 million years ago in gnathostomes (early vertebrates; jawed fishes) is well established, although debate over where the constituent components came from continues (Dreyfus, 2009). Rolff (2007) puts forward a multi-causal selective scenario of pressures, unique to the early vertebrates: coevolution with specialised parasites; increased metabolic rates, and genomic instability, that are claimed to be sufficient to select for the adaptive immune system. These pressures can be seen collectively as an increased pathogenic load: specialised parasites clearly; increased metabolic rate increasing feeding, and therefore pathogen exposure; genomic instability increasing the danger from within — an endogenous pathogenic threat.

A contrasting perspective considers that the benefits of extending the genome through managing a commensal or mutualistic flora are sufficient to select for a more specific immune response. Augmenting the innate immune system with sensors able to discriminate more accurately between bacteria that share common antigenic patterns and would trigger a more primitive innate-only immune system (McFall-Ngai, 2007).

Whilst immune systems did not stop evolving with the advent of the adaptive system, it marks the last major transition in our history. Perhaps, following Maynard Smith and Szathmary (1998),
who included sociocultural transmission as their final transition we should include the use of the same, in the forms of vaccination, medicines and epidemiology: a major transition in the evolution of immune systems.

### 2.5 Artificial Immune Systems

We have discussed the composition and evolutionary history of biological immune systems, but interest in their properties has not been restricted to the medical profession. In the same vein as earlier work on genetic algorithms and neural networks, researchers have noted that immune systems, in particular the adaptive immune system, appear to have computationally useful properties, exhibiting pattern-recognition, generalisation, learning, memory and adaptation, in a robust, scalable, decentralised, self-organising and distributed architecture.

Artificial Immune Systems (AIS) are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving. (de Castro and Timmis, 2002, p. 58)

Timmis and Andrews (2007) provide a thorough history of the origins of artificial immune systems, which we summarise here. The field has its roots in theoretical immunology, in models of idiotypic immune networks first suggested by Jerne a decade before (Jerne, 1974), and how such networks might maintain immunological memory (Perelson, 1989; Farmer et al., 1986; Varela et al., 1988). These early papers, particularly Farmer et al. (1986), described immunological models in terms of machine learning, which helped to spark interest from the computer science community.

Stephanie Forrest, working with Alan Perelson, initiated interest in applying immune system mechanisms to computer security and network intrusion detection, a significant theme for the community (Forrest et al., 1993, 1994; Hightower et al., 1996; Forrest et al., 1997; Hofmeyr and Forrest, 1999; Forrest and Hofmeyr, 2000). This route, from theoretical-immunology origins to a more engineering-focused perspective, is illustrative of a wider trend in the field: towards engineering applications and away from the immunology that inspired earlier researchers. This is seen by some as a problem to be addressed by the community, to return to a dialogue with immunologists rather than just taking inspiration from the literature (Timmis et al., 2008a).

Cohen (2007), in framing the immune system in computational terms, classified the activities in AIS into three areas: the literal school, metaphorical school and computational immunology. We will use Cohen’s categories (his descriptions in italics) to discuss the different approaches and how they fit with our own goals.
2.5.1 The literal school

AIS scientists of the literal school attempt to construct algorithmic systems in silico that can do what real immune systems do in vivo: protect computers, for example, from computer viruses by deploying algorithms designed to mimic receptors that discriminate between self and non-self, neutralize viruses with antibody-like agents, and so forth.

The literal school of AIS research attempts to construct algorithms in silico to solve analogous problems to those tackled by the biological system in vivo: to protect computer systems from viruses, worms and other intrusions through processes of recognition and classification (normal / anomalous rather than self / non-self). Whilst there is not a 1:1 mapping between elements in a computer immune system and the biological system – the substrates are very different – the underlying claim here is that the form (in terms of elements and their interactions) is very much responsible for the function of immunity. The difficulty comes in deciding at which level(s) of abstraction to extract the biological representation (i.e., the physico-chemical, cellular, and tissue-level views) – where is the computing occurring?

The kinds of threats a biological organism encounters, and that the immune system has evolved to cope with, develop in a different way to those encountered by a networked computer. Biological pathogens evolve by myopic ‘tinkering’, albeit in a massively parallel fashion, but blind to the wider reaches of the fitness landscape they are ascending. A well-planned and resourced attack on a computer system, however, has the benefit of a designer – an engineer who has the target system’s source code, and knows the weak points, i.e., they have the benefit of a map of the fitness landscape (or ‘attack surface’).

Finally, work initially motivated in the literal school is always open to the temptation to take a promising model or algorithm and drift away from the original biological inspiration if this will achieve an engineering goal, which leads to work similar to that of the metaphorical school.

2.5.2 The metaphorical school

AIS scientists of the metaphorical school, who comprise most of the AIS community, look to the immune system for inspiration; they do not try to mimic or simulate algorithm-like systems designed by evolution, but they aim to design new algorithms with the immune system in mind.

The metaphorical school does not aim to remain faithful to the in vivo implementation, but to build algorithms that work, inspired by abstractions of immune processes. Garrett (2005) gave a useful review of this approach, evaluating the principal algorithm classes (negative selection, clonal
selection, immune networks and danger model) to discern whether they are truly algorithmically
distinct from other approaches such as genetic algorithms and machine learning (merely being
biologically-inspired doesn’t count) and whether they are effective algorithms (competitive in
terms of quality of results or speed in computing them). An important reference to the ‘no free
lunch’ theorem is made: all algorithms are good at at least one problem, so for an algorithm to be
useful it must be distinct, and effective over a class of problems of practical interest (Wolpert and
Macready, 1995). Hart and Timmis (2008) take this line of argument further — claiming that
AIS has to be more than merely better or faster, it has to bring a unique approach, to distinguish
itself from other techniques to really contribute.

An area which has generated interest surrounds the Dendritic Cell Algorithm (and later develop-
ment, the deterministic DCA) (Greensmith et al., 2005, 2008b; Gu et al., 2013), which brings new
detection features into an AIS model – the damage associated danger signals from danger theory
(Matzinger, 2002). Danger signals are integrated with antigen detection signals to give context to
the non-self signatures collected by a population of simulated dendritic cells, this helps to classify
the anomalous non-self as dangerous or just unusual – a model based on the real dendritic cells
of the innate immune system, and carried out in conjunction with empirical work by wet-lab
immunologists (Greensmith et al., 2008a). In an example of the kind of critical analysis that
serves the field well, Stibor et al. (2009) performed an important dissection of the dendritic cell
algorithm and showed that it can be decomposed to a set of linear classifiers and cannot therefore
be considered distinct to other machine learning techniques. An example of convergent evolution
in algorithms, perhaps, but we learn something from the process of discovery.

Timmis et al. (2008b) outlined the main immunological theories and the corresponding algorithms
that drive AIS: negative selection, clonal selection, immune networks and the dendritic cell
algorithm. Most activity centres around learning and memory mechanisms (clonal selection and
immune networks) and negative selection for anomaly detection (changes in self). Reviews of
applications in this area include Dasgupta (1998); Dasgupta and Azeem (2007); Hart and Timmis

Cayzer et al. (2005) developed an interesting approach to efficiently covering an antigen shape
space. Bitstring receptors that match antigen are created through combining elements from
gene-libraries – borrowing the adaptive immune systems combinatoric pseudo-random VDJ
process. The contribution here is to note that the makeup of the elements in the gene-libraries
has a biasing effect on the receptors that are created. Adaptation of the gene library elements
is itself a form of meta-learning, contributing to the environmental representation performed
by the immune system, and ultimately to performance (Cayzer and Smith, 2006; Cayzer and
Sullivan, 2007). To give an example, the feature space for a text-classification problem might
use individual words (‘cat’, ‘watermelon’, ‘antidisestablishmentarianism’) as the features, but
in a truly evolved system these fragments themselves are under selection and should come to
Chapter 2 Background

represent the underlying compositional nature of the system being targeted (e.g. variable length \( n \)-grams).

A challenge for the AIS community, raised by Garrett (2005), is finding its niche, i.e., determining which real-world problems are most suitable for an immune-inspired approach (Hart and Timmis, 2008), and determining where this approach can contribute true algorithmic novelty to computer science. There is optimism about the future of the field as it matures (Greensmith et al., 2010), and increasing crossover between AIS and other bio-inspired paradigms such as swarm intelligence (Timmis et al., 2010). However, in a more recent review, Dasgupta et al. (2011) conclude that, while there have been an extensive number of AIS applications developed, the field is still limited by a lack of an exemplar, a ‘killer application’ for AIS.

2.5.3 Computational immunology

However, there is yet a third group of AIS scientists, who instead of exploiting the immune system to solve the challenges posed by computer sciences, aim to better understand immunity by developing computer models of an organism’s immune system.

The final group of researchers are focused on immunological goals rather than engineering, a perspective opposite to those of the other two schools. Computer models are employed to understand aspects of immune system mechanism and ontogeny, but still with a strong bias towards the proximate immune systems of humans and model animals used in \textit{in vivo} experiments. This school is perhaps better described as immunological systems biology (Germain et al., 2011), and the models range from the molecular level through to epidemiological, employing computers to mine the huge wealth of experimental and public-health data (Petrovsky and Brusic, 2002).

2.5.4 A fourth school?

The work in this thesis does not fall comfortably into any of the three schools Cohen describes. Rather we are asking functional evolutionary questions about the structure of the problem the immune system solves – and why such structure is maintained at all for the immune system to exploit. We are not aiming to protect computers, so the work is not in the literal school; we don’t plan to build new or improved algorithms to work the analogous problems of the metaphorical school; and we are not in the business of building computational models of the real immune system, or taking a systems biology approach.

We are not alone – when Cayzer et al. (2005) considers meta-learning, the evolution of the representation; when de Castro and Timmis (2002) propose a structure for engineering AIS
from representation to algorithms; when Stepney et al. (2005) discuss a meta-framework within which domain-independent AIS models should be built and assessed; Timmis et al. (2008a) outline a vision of immuno-engineering; when McEwan and Hart (2009) discuss representation in (artificial) immune systems; and when Hart et al. (2013) discuss distilling principles of immunology and the importance of minimising superfluous elements in the representation there appears to be a common theme – abstracting the necessary computational elements, i.e., those that operate on a given substrate, from the substrate — the contingent implementation details that are specific to a given problem domain, building things with organic chemistry or software modules.

To attempt to unify this school, which we identify here, we would claim that the aims are not necessarily to contribute to medical or therapeutic understanding of our immune systems, but to pare back the complexity implicit in biological immune systems, both proximate and ancestral, and understand them in a more abstract, strategic light. All living things ultimately deal in energy, matter and information (in a biosemiotic sense, it’s all symbols). This school is about understanding immune dynamics computationally (in a universal, rather than silicon-based sense), not towards any particular application, but in and of themselves. It is also about the methodology required to perform the separation of computational properties from specific implementations – recognising the enabling constraints in one problem domain with an observable solution, better to identify new domains where such a solution might be brought to bear.

This thesis contributes to this kind of ‘methodological glue’ – identifying enabling constraints (here, an ‘optimal’ level of structured epistasis) in evolved systems that give the problem a particular structure. Wolpert and Macready (1995) tell us that there is ‘no free lunch’, that there are no general purpose optimisation algorithms, and that all successful algorithms are attuned to a specific problem. If AIS practitioners want to apply their extracted algorithms to new domains, then it is imperative that we properly characterise the class of problems they are attuned to, what problem features enable this, and that we know how to recognise them.

To refer back to the lack of a ‘killer application’ for AIS, perhaps it it not a product, a specific biologically inspired representation / algorithm, but rather a process, a meta-framework for bringing ensembles of feature-selection, adaptive-representation and search algorithms to bear on any given problem.

2.6 Summary

We have covered a history of immunological discovery and thought, and a brief introduction to the components and activity of the human immune system. We discussed the transition from the self-non-self paradigm that ruled 20th century immunology to the danger theory model, marking
Chapter 2 Background

a shift towards a more integrated view of adaptive and innate immunity and the importance of highly conserved markers of pathogens and other ‘danger’.

These highly conserved signals provide useful information if you can interpret them, warranting a nod to biosemiotics, before we moved to evolutionary biology to ask how it is that some patterns can become highly conserved, and thereby facilitate immune recognition.

We used Tinbergen’s four questions about evolved traits to identify that our questions are on function — in asking not just about immune systems as-they-are, but as-they-must-be, if Gould’s tape of life could be played again (Gould, 1990). This type of question means abstracting beyond the single example we have of immune systems evolving on Earth in an organic substrate. Extracting the essential information processing from the messy organic substrate brings us to computational modelling and the field of Artificial Immune Systems.

We introduced the AIS field and used Cohen (2007)’s three schools to help review existing work, and orientate our own. We found that our questions do not sit naturally in any of the three existing schools, and argued that our work (and others’) belong in a fourth school. In this school the principle activity is abstracting computational principles from natural systems, but this is distinguished from the literal or metaphorical schools in that the goal is to better understand the process in and of itself, rather than to generate an algorithm tuned to some domain problem.
3 Materials and Methods

The preceding chapters introduced the motivation and direction of this thesis, and reviewed a selection of relevant literature on immunology, evolution and artificial immune systems. The purpose of this chapter is to return to the core research questions stated in Section 1.4 and restate them in the light of the technical material that we have covered since that point. We introduce the NK model in some technical detail, and discuss the choice of this framework to address our research questions. Finally, we state the specific, testable hypotheses that derive from the research questions and guide the experiments in the following chapter.

3.1 Conserved recognition and enduring patterns

We observe highly conserved Pattern-Recognition-Receptors (PRRs) in host organisms (notably, the Toll-Like-Receptor family – Section 2.2.2). Close receptor homologues exist in plants, fruit flies and humans, which tells us that these PRRs have contributed high fitness to these host lineages for at least 500 million years, since the last common ancestor of insects and animals (Beutler and Poltorak, 2001; Leulier and Lemaitre, 2008), and perhaps as long as a billion years, since the last common ancestor with plants (Ronald and Beutler, 2010).

Two possible evolutionary explanations could account for the apparent homology we see today in PRRs: either they represent divergent (but constrained) evolution from an ancient common
eukaryotic ancestor of plants, insects and animals, or, the apparently analogous mechanisms we observe today actually represent convergent evolution and reflect inherent constraints guiding multiple evolutionary routes to the same solution (Ausubel, 2005).

Either explanation tells us that the underlying fitness landscape has significant and enduring structure that causes such phylogenetic paths to be taken. In the divergent evolutionary story, we start from the same point, but are so constrained by the landscape that the numerous routes explored never stray far from one another. Conversely, in the convergent evolutionary story, the landscape has such a large and deep basin of attraction that all (initially diverse) proto-immune sensors converge – arriving at analogues close enough to mistake for homologues. We arrive at the same conclusion – the number of ways to build innate immune sensors is highly constrained for some reason.

Highly conserved receptors are highly conserved only as long as they contribute high fitness to the host organism’s immune system. For this to be true, the signals that they detect must be similarly conserved. Pathogens would clearly have an advantage if they were able to throw off easily recognised coats, but are apparently energetically and evolutionarily unable to do so\(^1\) – the recognised parts of their genetic architecture are deeply entrenched.

Charles Janeway originally made the case for pattern-recognition-receptors in his landmark paper “Approaching the Asymptote” in 1989. He proposed that genetically encoded, not adaptively learned, microbial pattern recognition sensors had an important role in boot-strapping the immune response – an idea that helped synthesise an integrated view of the innate and adaptive immune systems\(^2\).

> “The pattern recognised should be the product of a complex and critical enzymology in the microorganism. Complex cell wall carbohydrates or lipopolysaccharides are likely ligands. Such structures ... would require multiple mutations and major evolutionary shifts on the part of the microorganism to evade recognition. Presumably, over evolutionary time, the receptors would reach a standoff with the microorganism at a stage where the possibilities of change in the microorganism have been severely reduced.”

A functional ‘why’ explanation for the conserved immune receptors will be found in the evolutionary constraints that prevent the targets from freely changing, even when detected. The

---

\(^1\)Although they do have many strategies for evading detection (Kugelberg, 2013): subverting the signal when detected, or otherwise mitigating the subsequent immune response (Arpaia and Barton, 2013), deploying these strategies always comes at a fitness cost against the wild type.

\(^2\)Janeway (1989) also helped explain why adjuvants were required for successful vaccination: the pathogen-associated molecules (e.g., lipopolysaccharide) in the adjuvant were required to mimic a microbial infection, and ‘trick’ the immune system into responding to the antigen (Medzhitov, 2009).
structure in the recognition problem that immune systems solve derives from these evolutionary constraints. To address this kind of question, we need to develop some insight into how evolutionary constraints shape fitness landscapes.

3.2 Epistasis and the NK model

The term ‘epistasis’ was first coined by Bateson (1909) to describe the phenomenon wherein the presence of an allele at one locus of a genome could override the known effect of an allele at a different locus – what he termed a masking effect, which demonstrated an interaction between the two loci (Cordell, 2002). Epistatic interactions between alternative alleles at two given loci on the genome create the possibility of multiple conflicting solutions. For example, either of two alleles individually are deleterious (Ab or aB), but both (AB), or neither (ab) contribute a high fitness score\(^3\). Fitness depends on well-adapted cohorts of alleles, and less on the presence of individual specific alleles – making analysis of such systems statistically more difficult, and frustrating headline writers who enjoy writing about the discovery of the ‘gene for X’.

Understanding epistasis is fundamentally important to understanding the dynamics of the complex regulatory networks that underlie all life (Phillips, 2008), but also in working with non-biological multi-optimisation problems that have interactions between their component variables.

To characterise the link between increasing levels of epistasis and evolvability – that is, the ability for a simple evolutionary process to discover ‘good’ solutions in a search space, Kauffman and Weinberger (1989) created the NK model of fitness landscapes\(^4\). A fitness landscape takes the phase-space, that is all possible configurations of variables, and adds one extra dimension, a fitness score. When Wright (1932) introduced the fitness landscape he restricted himself to two dimensions plus height. This turned a conceptually difficult mathematical idea about population genetics into a visually accessible metaphor, with populations roaming the valleys and hills raised up by the interaction of the variables, as happens under epistasis.

In Kauffman’s model, the parameter \(N\) controls the number of variables, or dimensions, and \(K\) determines the degree of epistasis operating on each locus. The fitness associated with a particular genotype (i.e., the height associated with a particular point on the fitness landscape) is assessed by combining the fitness contributions of the binary alleles at each of its \(N\) loci. The fitness contribution of a locus, \(i\), is determined by the allele at \(i\) and the alleles present at \(K\) additional loci (see Figure 3.1), specified by the epistatic network. For each unique combination of \(K + 1\) alleles, a unique, but randomly determined fitness contribution is assigned. By considering the

---

\(^3\)Here corresponding to the XOR boolean operator

\(^4\)The NK model has its ancestry in the field of spin-glass models from statistical physics. In these models the spins (up or down) of \(N\) atoms, each affected by \(K\) others, must be arranged so as to minimise the energy of the whole collection of atoms.
statistical properties of ensembles of NK landscapes, the generic influence of epistasis can be assessed.

Kauffman was able to show that the ‘ruggedness’ of the landscape increases with increasing $K$. Epistatic interactions represent constraints that frustrate evolutionary search to some degree, so as $K$ increases, it becomes more difficult to resolve the conflicting constraints. This progression is broadly characterised as follows:

$K = 0$
When there is no epistasis, no constraints to satisfy, all $N$ genes that contribute to system-level fitness do so independently of one another. The landscape is smooth, with the fitness of adjacent genotypes being highly correlated as a consequence of sharing $N - 1$ fitness components. An adaptive walk originating at any point on such a landscape will reach a single unique optimum. This optimisation problem, assuming binary alleles, is solvable in linear time: every step on an adaptive walk reduces the distance to the global optimum by one as the allele at a single locus is mutated to the fitter variant. In the worst case all $N$ loci need to be mutated, giving a mean walk length of $\frac{N}{2}$.

$K \approx 2$
When $K$ is small, but non-zero, epistatic interactions increase the pool of possible fitness contributions for each locus (for binary alleles, from 2 to $2^{(K + 1)}$) and improve the possibility that a high fitness value exists, and can be discovered. Some interactions between components allow for much richer system-level behaviour (exploiting synergies between components) while keeping the conflicting constraints (anti-synergies) at a manageable level for a search algorithm to handle, allowing it to find optimal configurations in a reasonable time. This leads to a fitness landscape with multiple peaks at higher fitness values than at $K = 0$. Also, these higher peaks are clustered together, creating the ‘Massif Central’, a global structure in the landscape.

$K \approx 4$
As $K$ increases, the number of epistatic interactions make it increasingly difficult to optimise any single locus without negatively affecting the fitness contribution from other loci. Due to such competing interactions, ‘frustration’ in the system increases, giving rise to what Kauffman called the ‘complexity catastrophe’ – unresolved constraints further buckle the landscape into increasing ruggedness, altering the average distance between local optima, the correlation amongst locally optimal genotypes, and the creating increasing numbers of local optima with average fitness.

$K = (N - 1)$
When $K$ is maximal, the epistatic dependency graph is fully connected. A mutation at any locus has the effect of changing the fitness contribution of the alleles at all other loci.
Consequently, there is no correlation between the fitness of neighbouring genotypes, and the landscape is maximally rugged. A large proportion \( \frac{1}{\sqrt{N}} \) of genotypes are now local optima, and adaptive walks tend to stall after \( \ln(N - 1) \) steps Altenberg (1997). The only strategy that can guarantee to find the global optimum is exhaustive search and, in this case, the problem is NP-complete (Weinberger, 1996).

![Diagram](image)

Figure 3.1: Calculating fitness contributions for an \( N = 3, K = 2 \) fitness landscape example. 1) The dependency graph is shown as both an adjacency matrix \( A \) (all values except the diagonal set to 1 — the maximal \( N = K - 1 \) case), and as a network diagram (nodes in green). 2) The table shows the fitness contributions for each allele in each combination of its relevant \( K = 2 \) neighbours. 3) A three-dimensional boolean hypercube, with a bitstring, representing a binary genome at each of the \( 2^3 \) vertices. Each of the three neighbours of each vertex is a single 1-bit mutation away. Note that two nodes are local optima (red), and that a non-greedy adaptive walk starting at 010 could reach either of them. Adapted from Kauffman (1993).

In the two most frequently explored forms of the model, the \( K \) loci that epistatically influence a particular locus, \( i \), may either be randomly located on the genome, or may be the \( K \) nearest neighbour loci of \( i \). In both cases, every gene influences the fitness contribution of (on average) the same number of other genes, ensuring that genes are equipotent in their contribution to genotypic epistasis.

Within this simple framework, there are still many elements to explore, and there is a great deal of literature on the effects of varying these parameters.

Kimura (1983) argued that fitness landscapes contain fitness-neutral plateaus – where an evolving population spends time undergoing nonadaptive neutral mutation. Geard et al. (2002) provides a useful comparison of two models used to explore the effects of neutrality on evolvability on
neutral landscapes: Barnett (1998) introduces the $NKp$ model with new parameter, $0 < p < 1$, which controls the probability that some allelic combinations have their random fitness value clamped to zero, giving a neutral step; Newman and Engelhardt (1998) quantise the fitness values to discrete intervals, giving rise to a ‘terraced’ fitness landscape. Variations on these techniques have been developed by others (Correia, 2012; Smith et al., 2002; Verel et al., 2003).

Soberingly, Kauffman (2012) also reminds us that there is a “profound limitation” to the mathematical and computational tools with which we might try to model biological dependencies. This is because the state space of most of our models is pre-defined and therefore could be pre-computed, whereas the powerful thing about real-world state spaces is that they are forever expanding into the “adjacent possible”. As Kauffman puts it, there is no algorithm that will let you spell out all the possible uses for a screwdriver.

However, we need to try to model the emergence of biological dependencies if we are going to make sense of evolution in general and the evolution of immune systems in particular. As Jacob (1977) argues, the big steps in evolution required the acquisition of new information, but specialisation and diversification are essentially just fiddling with the details. Kauffman finds in his simple models that a low level of epistasis, of interconnectivity between traits, actually improves the fitness of solutions found over the baseline case with no interactions between traits at all. Yan et al. (2010) look at modularity, inter- and intra-module dependencies, and the evolution of modularity during the exponential growth of the Debian GNU/Linux operating system. Their conclusion is that conflicts between modules actually has a positive effect as it allows exclusion of incompatible modules and speeds selection.

The lesson here is that, in a complex system, the interactions between components mean that changes to one small component may have effects that cascade through the system. The structure of the network of interconnections is a source of non-linearity in the behaviour of the whole system. Such non-linear systems have the property that small changes in input can be the cause of large changes in behaviour: i.e., amplification. Conversely, we also see robustness — the system manages to maintain its state despite large perturbations to its inputs.

When we are looking at living biological systems, the structure of these networks of connections, between genes, proteins, cells or even social individuals, have evolved over time so as to confer ‘fitness’ on the reproductive units that create them. If a network is wired up so as to accept inputs and produce outputs that improve the likelihood of that network existing again in the future, it stands a competitive chance of doing so. From the first auto-catalytic sets of chemicals that managed to replicate themselves over time, the structures we see are those that best model the world around them. In other words they embody structures that process information in a way that responds appropriately to the inputs from the world, and produce outputs that include more copies of themselves that are adapted to the world. Being adapted to the world includes coping with change, so allowing mutation and other methods that accumulate ‘useful’ information
into a lineage, such as sexual recombination, are themselves improving models of the world. Maintaining information can be expensive, so going one step further and learning patterns in the world is a good idea — this is a form of data compression, or exploiting the low entropy of ecosystems.

### 3.3 Research questions, hypotheses and experiments

Here we revisit the three specific research questions detailed in Section 1.4. In light of the literature review and discussion, we give specific, testable hypotheses, and how these will be supported (or rejected) by our planned experimental results.

#### 3.3.1 RQ1: Is there a class of problem space where evolutionary lock-in can provide slower evolving hosts sufficient information to identify faster evolving pathogens?

Our hypothesis: A class of fitness landscapes (some subset of all problem-spaces) representable in the $NK$ model, exists that has broad adaptive value (i.e., is well suited to evolutionary search) and exhibits evolutionary lock-in of some variables.

Our assumptions:

- The problem-spaces representable in the $NK$ model are a reasonable proxy for larger problem spaces (e.g., with differing numbers of alleles per locus, different neighbourhood operators, etc.).
- locked-in alleles represent a sufficient signal for immune recognition by a slower evolving host or a faster adapting pathogen.

We will retain the $NK$ framework in all aspects for comparison to existing published results, and only modify the epistatic network topology. This includes keeping a uniform $K$ in-degree for all loci. Existing work in the literature has found the $NK$ model to be a valuable means of gaining insight into evolutionary dynamics and the effects of epistasis, despite (because of?) its simple set-up. We want to ask a closely related question about structured epistasis and measure the comparative effects.

We have discussed the issue of lock-in and host / pathogen recognition elsewhere, but it is worth making this assumption explicit — there are certainly situations we could imagine where the only locked-in features are universal, and do not help to differentiate between host and pathogen traits at all.
This research question is an umbrella for the two following questions – if we can find supporting evidence for them both, that our biologically-inspired epistatic topology generates lock-in, and that it has other evolutionary advantages as well, then we will have discovered a useful property to classify this set of problem-spaces. Namely that problems with a specific kind and level of structured epistasis (interdependency between variables) feature entrenched structures (a level of conserved redundancy, or low entropy) that immune algorithms can exploit.

3.3.2 RQ2: Do more ‘biologically-inspired’ epistatic graphs exhibit greater ‘lock-in’?

Our hypothesis: when we change the NK model epistatic graph topology to have a scale-free out-degree distribution, there will be a measurable increase in the number of alleles that fixate early in adaptive walks. This will also be measurable as a conserved set of alleles (that fixed early, and became locked-in) in the set of local optima discovered on the adaptive walks. There will be a direct relationship between the out-degree of a locus and the conservation of a particular allele at that locus in the set of local optima discovered.

Our assumptions:

- An appropriate baseline for comparison will be the published results for Kauffman’s local and random epistatic topologies (which we will reproduce as part of our implementation validation).
- The inequality we introduce, making some loci more influential than others will be the cause of lock-in.
- Qualitative differences will be measurable in the sizes of network / landscape that it is feasible to simulate.
- Lock-in can be adequately measured by recording the out-degree of mutants during adaptive walks.

3.3.3 RQ3: Can a ‘biologically-inspired’ epistatic graph improve evolvability?

Our hypothesis: when we change the NK model epistatic graph topology to have a scale-free out-degree distribution, there will be quantitative improvements in the evolvability of the fitness landscape.

Improvements in evolvability will be measured by the fitness of local optima reached, the length of adaptive walks, and comparison with the measures of basin size and global structure well defined in the literature.
Our assumptions:

- The scale-free distribution observed in natural networks is selected for, and contributes to the behaviour of the system, rather than being a coincidental artefact.

- An appropriate baseline for comparison will be the published results for Kauffman’s local and random epistatic topologies (which we will reproduce as part of our implementation validation).

- The changes that we introduce to the fitness landscapes through scale-free epistasis will be captured by the existing measurements used to assess fitness landscapes during adaptive walks.

- Qualitative differences will be measurable in the sizes of network / landscape that it is feasible for us to simulate.

- Greater degrees of inequality (a more skewed out-degree distribution) will provide a more significant gradient for an adaptive process to work with. In this sense, the additional structure in the epistatic network is an enabling constraint.

3.4 Summary

This chapter returned to the idea of functional questions about the immune system, and why asking questions about the forces (epistasis) that shape evolution are the best way of understanding immune system development in a general way, transcending the particular biological instance.

We justified the choice of Kauffman’s NK model from the literature as a widely used and validated, computationally tractable and easily extended model for investigating the effects of epistasis on evolutionary dynamics. Some detail on the working of the model was provided, and of the results we should expect to see to validate our own implementation of the model.

Finally, we returned to the three specific research questions from Section 1.4, and reformulated them as testable hypotheses that can be put to our implementation of the NK model. We described the kinds of measurements that will be required to test the hypotheses, as well as stating our assumptions about these experiments and the kinds of results we expect to see.
Scale-free epistasis in the $NK$ model

Recent advances in our understanding of natural genomes are beginning to reveal patterns in genomic organisation (Oltvai et al., 2000; Barabási and Oltvai, 2004; Segrè et al., 2004; Hu et al., 2013). In particular, the epistatic networks that describe the manner in which genetically specified proteins interact with each other during cell metabolism have been shown to exhibit topologies that are scale-free in their degree distribution (Maslov and Sneppen, 2002; Fernández, 2007). In such networks, while the vast majority of proteins are involved in only a small number of protein-protein interactions, a few proteins are highly influential (Barabási et al., 1999).

One of Kauffmans’s most remarkable results from his original work on the $NK$ model was the discovery of the ‘Massif Central’ at low, but non-zero levels of epistasis ($K \approx 2$). A simple algorithm can exploit this structure to find good solutions in the space, as they tend to be clustered together. However, at higher levels of epistasis, the increasing number of conflicting constraints that any search algorithm must resolve cause a ‘complexity catastrophe’ overwhelming the ability of evolutionary search (or any algorithm) to find good solutions (Kauffman, 1993).

In this chapter, inspired by natural networks, we bring a scale-free epistatic topology to the $NK$ model. Our results show we can mitigate the complexity catastrophe by replacing Kauffman’s uniform or random topology with one that has a scale-free out-degree distribution. All the original framework of the $NK$ model remains, so our new results can be compared to those in the literature. We maintain the standard in-degree per-locus at $K$, but now allocate edges in the epistatic graph
using a rich-get-richer method, so the out-degree distribution approximates a power law with exponent \( \alpha \) – a new parameter that controls the extent to which some loci are more influential than others\(^1\).

Our results show that fitness landscapes linked to such epistatic networks contain clusters of high local optima with large basins of attraction. Comparing these new landscapes to those using the previous uniform or random epistatic topologies, the interesting properties (the Massif Central) are accentuated, and persist at higher levels of epistasis.

Where there is inequality of influence (high \( \alpha \), low to medium \( K \)), mutating the most influential loci substantially changes overall fitness. Consequently, mutation events at high-degree loci tend to happen early in adaptive walks, when a large change in fitness is more likely to be beneficial. They happen less often as fitness increases towards the end of adaptive walks, when any significant change in fitness is usually detrimental.

We expected that having high-degree loci fixing early (‘lock-in’) in adaptive walks would result in certain alleles being highly conserved throughout the high optima on a landscape, thereby contributing to the clustering of high optima (the Massif Central). However, we found that while the Massif Central certainly persists, this global structure is explained by highly conserved groups of coadapted alleles, rather than individually identifiable conserved alleles.

A second round of experiments in which we manipulated the epistatic network topology to have particular assortativity (where high-degree nodes preferentially attach to other high-degree nodes, and visa versa) were inconclusive. We observed no additional benefit to tuning epistatic networks to be dissortative (as many natural networks are), and only a slight penalty when tuning to be assortative.

### 4.1 Scale-free Networks and Assortativity

Scale-free degree distributions have been discovered to characterise connectivity in a wide variety of systems, from gene regulatory networks to scientific citation networks (Barabási et al., 1999; Rzhetsky and Gomez, 2001; Gisiger, 2001; Albert, 2005; Barabási et al., 2002; Barabási, 2003). In each case, the frequency with which network nodes exhibit degree \( k \) is proportional to \( k^{−\gamma} \), where \( \gamma > 1 \) (Barabási, 2003). Scale-free networks of this kind may be grown via a process of ‘preferential attachment’ (Barabási et al., 1999; Newman, 2001; Caldarelli et al., 2002; Eisenberg and Levanon, 2003). Under such a scheme, nodes are added sequentially to an initial small graph. Upon being added to the graph, each node is allocated a number of edges linking it to existing

---

\(^{1}\)Kauffman himself mentions briefly a variant of the model in which some genes are more influential than others (Kauffman, 1993, pp78). We develop this idea and explore the implications of systematically manipulating the extent to which there is a particular scale-free non-uniformity in the magnitude of influence exerted by each gene on the fitness contribution of the remainder of the genome.
nodes, where the probability of adding an edge to an existing node of degree \( k \) is proportional to \( k^\alpha \). Here, \( \alpha \) is a model parameter governing the strength of preferential attachment.

Networks with a scale-free topology have some distinct properties:

**Self similarity at different scales:** properties of local areas of the network are echoed in the whole.

**The small-world phenomenon:** shortest paths between any pair of nodes are remarkably short (Watts and Strogatz, 1998; Albert et al., 1999; Lazer and Friedman, 2005; Giacobini et al., 2006).

**Robust to random failure:** removal of nodes at random has little effect on network structure. However they are vulnerable to attacks that target the highly connected hubs (Albert et al., 2000; Barabási, 2003).

A network’s assortativity is the extent to which a pair of nodes connected by a network edge tend to resemble one another. In an *assortative* network two connected nodes will tend to share properties. In a *dissortative* network the opposite is true, and connected nodes will tend to be dissimilar. Here similarity will be measured in terms of *out-degree*, i.e., nodes \( i \) and \( j \) will be judged to be similar to the extent that they influence the fitness contributions of the same number of other nodes. A network’s assortativity will consequently be measured in terms of its *degree-degree correlation*, \( r \), which ranges from \( r = 1 \) (maximally assortative) through \( r = 0 \) (uncorrelated) to \( r = -1 \) (maximally dissortative).

While a random graph built via an Erdős-Rényi process has zero degree correlation, many real-world networks are either positively or negatively assortative (Newman, 2002). In particular, many biological networks are significantly dissortative, including some protein-protein networks — see Table 4.1 for examples.

### 4.2 NK\( \alpha \) implementation

An extensible \( NK \) model was implemented using a variation on the hashing method described by Altenberg (Altenberg, 1994, 1997), and validated against published data from several sources for the Kauffman local and random variants (Kauffman, 1989; Weinberger, 1991; Kauffman, 1993, 1995; Altenberg, 1997). We used a high performance hashing algorithm\(^2\) proven against funnelling effects (Jenkins, 1997) whereby correlation in the input bitstrings can result in correlation in the random real-valued outputs, i.e., there would be a danger that the random fitness values would not be completely random. Fitness is calculated as:

\(^2\)Austin Appleby’s MurmurHash 2.0 [http://sites.google.com/site/murmurhash/]
Table 4.1: Size $n$ and assortativity coefficient $r$ for various networks, taken from Newman (2002), see reference for details.

<table>
<thead>
<tr>
<th>Network</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics coauthorship</td>
<td>52909</td>
<td>0.363</td>
</tr>
<tr>
<td>Biology coauthorship</td>
<td>152025</td>
<td>0.127</td>
</tr>
<tr>
<td>Mathematics coauthorship</td>
<td>253339</td>
<td>0.120</td>
</tr>
<tr>
<td>Film actor collaborations</td>
<td>449913</td>
<td>0.208</td>
</tr>
<tr>
<td>Company directors</td>
<td>7673</td>
<td>0.276</td>
</tr>
<tr>
<td>Internet</td>
<td>10697</td>
<td>0.189</td>
</tr>
<tr>
<td>World-Wide Web</td>
<td>269504</td>
<td>-0.065</td>
</tr>
<tr>
<td>Protein interactions</td>
<td>2115</td>
<td>-0.156</td>
</tr>
<tr>
<td>Neural network</td>
<td>307</td>
<td>0.163</td>
</tr>
<tr>
<td>Marine food web</td>
<td>134</td>
<td>-0.247</td>
</tr>
<tr>
<td>Freshwater food web</td>
<td>92</td>
<td>-0.276</td>
</tr>
</tbody>
</table>

Random graph

<table>
<thead>
<tr>
<th>Callaway et al. (2001)</th>
<th>$\delta/(1+2\delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert and Barabási (2001)</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
f(x) = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \quad (4.1)
\]

\[
f_i(x) =hash(A_i \land x, s_i)\quad (4.2)
\]

Where $x$ is the genome bitstring (length $N$), $A$ is the adjacency matrix of the dependency graph, $\land$ is the bitwise / logical AND operator, and $s$ is a vector, length $N$, containing unique seed values for each row of $A$, usually initialised as the concatenated bits of the row number $i$ and the system-wide seed.

The network of epistatic interactions between loci was represented as an $N \times N$ Boolean matrix, $A$, with $A_{ij} = 1$ iff locus $i$ influences the fitness contribution of locus $j$. Since each locus always contributes to its own fitness contribution, $A_{ii} = 1 \ \forall i$. Furthermore, $\sum_i A_{ij} = K + 1 \ \forall j$, since each row of $A$ contains $K$ entries in addition to the self-connection, corresponding to $j$’s incoming edges. By contrast, the sum of each column of $A$ corresponds to the out-degree of each locus, which, in general, may be free to vary such that $1 \leq \sum_j A_{ij} \leq N$. Under all schemes considered here $\sum_{i,j} A_{ij} = N(K+1)$, i.e., the total number of edges in the network is conserved.

Kauffman’s original $NK$ model employed two schemes for allocating the epistatic links: local (also called adjacent) or random. In the former, each locus is influenced by its $K$ nearest neighbours, giving rise to an epistatic network with a ring-lattice topology (see Figure 4.1). In the latter, for each locus, $K$ unique influential loci are chosen at random, giving rise to a random graph
topology (see Figure 4.1). Under the local scheme both in- and out-degree are uniform, whereas under the random scheme in-degree is uniform, but the out-degree is a Poisson distribution with a mean of $K$ (Newman et al., 2001).

![Network Diagrams](image)

Figure 4.1: The adjacency matrix and network diagram for each of the epistatic network initialisation algorithms: local, random, scale-free, scale-free assortative and scale-free dissortative. In this diagram only, $N = 32$, $K = 8$, and $\alpha = 1.8$. For the assortative network, $r = 0.2$ and for the dissortative network $r = -0.5$. The colour of the network nodes represents the in-degree of that node, where darker is higher. Network plots all use the Hu (2005) proportional layout algorithm in order to emphasise structural differences between their topologies.

Here, we introduce NK$\alpha$, a variant of the NK model that employs a scale-free epistatic topology, parametrised by a single exponent, $\alpha$. As before, the network contains $N(K + 1)$ edges, $N$ of which are self-connections, and each locus has the same in-degree ($K + 1$). However, the out-degree distribution approximates a power-law as a consequence of the following preferential attachment growth process.

Initially each locus is connected only to itself, giving a degree of 1. Subsequently, we perform $K$ passes through the list of loci. Each pass visits each locus once in random order. On each visit, the visited locus is assigned one incoming edge from a random locus, $i$, chosen with probability $\propto (k_i)^\alpha$, where $k_i$ is the out-degree of locus $i$ and is updated after each visit, and the magnitude of $\alpha$ determines the strength of preferential attachment. This process assigns a total of $N(K + 1)$ edges with a power-law like degree distribution, save that a ceiling threshold exists: no locus can have more than $N$ connections (including its self-connection). With sufficiently high $K$ or $\alpha$, some loci will attract the maximum $N$ connections deforming the power-law curve. When $\alpha = 0$, the resulting epistatic matrix is equivalent to the random map explored in the original model. Where $\alpha > 0$, increasingly skewed degree-distributions are generated, conferring increasing influence on a minority of loci (see Figure 4.1).
Chapter 4 Scale-free epistasis in the NK model

Figure 4.2: Untuned assortativity values for epistatic networks generated using the $NK\alpha$ method before any assortative edge-swapping is performed. Solid line indicates the mean value of $r$, shaded ribbon the standard deviation. $\alpha$ increases the inequality between loci. As inequality rises, matching two loci with the same out-degree is less likely: assortativity falls. This is confirmed by the downward trend as $\alpha$ increases. Higher values of $K$ are less affected, as there are more edges to be distributed and a ceiling effect (no locus can have more than $N$ edges), so inequality remains lower even at high $\alpha$.

4.2.1 Adding assortativity $NK\alpha r$

In addition to exhibiting a characteristic degree distribution, networks generated via the process described above will also vary in their assortativity, some being more assortative and some less so (see Figure 4.2). We will sometimes refer to the ensemble of networks generated for any particular value of $\alpha$ as the ‘raw’ ensemble to distinguish it from the sub-set of networks that also satisfy a specific value of degree-degree correlation. To sample such a sub-set we employ the following algorithm.

To generate a network with parameters $N$, $K$, $\alpha$, and $-1 \leq r \leq 1$, we begin by creating a raw $NK\alpha$ network as per the description above. We measure the network’s assortativity $r'$. If $|r - r'| < \epsilon$ then we are close enough to the target assortativity and we can stop, otherwise we need to alter the network in a way that maintains both the in-degree and out-degree distributions, but changes the out-degree to out-degree correlation. To do this we take a pair of edges, $(i, j)$ and $(x, y)$, that involve four unique nodes, i.e., $i \neq j \neq x \neq y$. We remove the edges $(i, j)$ and $(x, y)$ and add two new edges $(i, y)$ and $(x, j)$ (neither of which can already exist in the network). By doing this we change which nodes $i$ and $x$ influence, without changing the number of edges in the network or the in-degree or out-degree of any of the four nodes (Figure 4.3 illustrates this swapping...
procedure). Figure 4.4 shows an example raw NKα graph, and then the same graph retuned to be assortative, and dissortative.

![Figure 4.3: Edge-swapping procedure to generate networks with desired assortativity r.](image)

![Figure 4.4: Epistatic maps for N = 32 with K = 8 and α = 1.5. (a) raw map, (b) rewired to r = 0.5, and (c) rewired to r = −0.5.](image)

In principle, by implementing a series of randomly chosen edge-swaps that each decrease |r − r'| by some amount, this process can generate networks with arbitrary degree-degree correlation for any values of N, K and α. However, in practice networks with extreme values of r can be very difficult to construct, especially for high values of α. For the results presented below, each sample of 100 networks for some combination of N, K, α and r was generated by rewiring a maximum of 1000 raw networks for up to 20,000 steps each, with ε = 0.01. We estimate the difficulty of generating a particular sample of networks by recording how many rewiring attempts were required (see Figure 4.5).

As α increases, deforming the distribution of the out-degree, the range of possible r is curtailed. For instance, with low K, high α you might get one very high out-degree node, and the rest all degree 1 (themselves). It is impossible to rewire this to anything but high assortativity, because all the nodes are very similar, and all of them affect themselves — a circular argument. Indeed, looking at Figure 4.5, we can see that as K increases there are more edges to allocate, and the
ceiling effect that no node can have more than out-degree of $N$ means that it is increasingly difficult to reconfigure to find extremes of assortative or dissortative topologies. Because of the choices of $K$, in the case of $\alpha = 0$ (equivalent to a Kauffman random topology) we see that it is easy to find configurations that satisfy the range of $r$. It is hard to find configurations with extreme $r$ partly due to the ceiling condition of out-degree $max = N$. As $K$ increases, or $\alpha$ increases, the algorithm creates skewed distributions where nodes have no ‘partners’. A dissortative graph has a low correlation between the out-degree of a given node and that of its neighbours. As $K$ increases, the potential diversity in out-degree is reduced as all nodes’ out-degree values approach the ceiling. Also, as $\alpha$ increases, the distribution of edges is extremely skewed, attaching those lucky first few nodes to be randomly chosen in a ‘rich-get-richer’ scenario, and then butting up against the same ceiling effect as nodes are saturated with connections, when out-degree $= N$.

When $N$ is very small (e.g. 4) the assortativity ($r$) tends to be negative (e.g., $N = 4$, across various $K$ and $\alpha$, mean $r = -0.6$, std. dev. = 0.3), and ranges from almost -1 to zero. When $N$ is larger, the contribution of any one locus ($\frac{1}{N}$) is reduced, and $r$ tends towards zero ($N = 96$, mean $r = -0.05$, std. dev. = 0.05). Whilst the $\alpha$ parameter does have an effect on $r$, (magnitude of $r$ is more negative as $\alpha$ is more positive), as $N$ increases, $r$ tends towards zero. (Two samples of $N = 20000$, $K = 8$, and $\alpha = 2$ gave $r = -0.00740258$ and -0.00490836.)

---

Figure 4.5: Each panel shows the mean error ($\|r - r'\|$) remaining, after search terminated, is shown for the values of $K$ and target $r$ for a given $\alpha$. The region in each panel with no colour shows the values of $K$ and $r$ where the error was under threshold $\varepsilon = 0.01$, i.e., was considered a success.
4.3 Results

For the model introduced here, a tuple \((N, K, \alpha, r)\) specifies an ensemble of landscapes that we sample and evaluate as below. In addition to sampling the fitness distribution over each landscape as a whole by evaluating 10,000 points at random on each landscape, we perform a number of adaptive walks across landscapes. At each step of an adaptive walk, a hill-climber calculates the fitness of all \(N\) single bit mutation neighbours of the current genotype, and selects one of the fitter neighbours at random to move to. If no fitter neighbour exists, then the hill-climber has reached a local optimum, and terminates. By undertaking multiple independent walks on the same landscape, an assay of available local optima can be compiled. Additionally, the length of adaptive walks is an indicator of a landscape’s ‘ruggedness’.

Unless otherwise stated, genotype length is held constant with \(N = 96\), \(K \in \{0, 1, 2, 4, 8, 16, 32, 64, 95\}\), and \(\alpha \in \{0.0, 0.5, 1.0, 1.5, 2.0, 2.5\}\). By ‘a full range of landscapes’ we will mean all combinations of \(N\) and \(K\) for the local and random variants of the original \(NK\) model, and all combinations of \(N, K\) and \(\alpha\) for the \(NK\alpha\) variant. For the majority of results presented, the data is an aggregation of results from 100,000 adaptive walks on 100 unique landscapes generated for those parameters.

Figure 4.6 shows the manner in which the distribution of adaptive walk lengths varies with \(K\) for the landscapes considered here. While, in general, walk length decreases with \(K\), it is also apparent that walks tend to be longer as \(\alpha\) increases (with a caveat that there is a ceiling to this influence of \(\alpha\), see below.). Moreover, while the distribution of walk lengths as \(K\) increases narrows for the random and local case, longer walks exist for \(NK\alpha\) landscapes for much higher values of \(K\), even as many as \(N/2\) steps at \(K = 64\).

Figure 4.7 demonstrates that increasing \(\alpha\) has an interesting effect on the distribution of optima fitness in comparison to the random and local variants. At lower values of \(K\) (\(K < 8\)), the local and random (and low \(\alpha\)) variant have the edge, but as the complexity catastrophe takes over at \(K \approx 16\) the higher \(\alpha\) scale-free topologies reach higher optima overall. As the sample sizes in these cases are very large (each distribution consists of 100,000 walks on 100 landscapes) the law of large numbers comes into play and we can be confident that visually distinct distributions are indeed different, without making assumptions about the normality of the underlying fitness distribution itself.
To confirm our conclusions from the data we ran the non-parametric Mann-Whitney’s U test (Mann and Whitney, 1947) on the distributions of optima height for various values of \( K \). The alternative hypothesis for these tests was that the distribution of optima found by the random epistasis distribution (our baseline) was greater than the distribution from the scale-free epistatic data. The resulting \( p \) values (due to the large sample size) were either 1 or at the numerical limit \( \approx 0 \). The epistatic threshold, when a scale-free topology finds higher optima than the random topology, is at \( K = 8 \) until \( \alpha \geq 2.0 \) when it rises to \( K = 16 \).

To summarise, when it comes to optima height, a scale-free topology, for all tested values of \( \alpha \), does not improve upon the existing local and random schemes unless epistasis is high (\( K > 8 \)).

We note that the influence of \( \alpha \) asymptotes (as measured by walk length or optima height in figures 4.6 and 4.7). The effect on the distributions of either measure is greater from \( \alpha = 1.0 \rightarrow 1.5 \) than it is from \( \alpha = 2.0 \rightarrow 2.5 \). Why should this be? Recall that increasing \( \alpha \) makes the epistatic out-degree distribution more unequal, i.e., stronger preferential attachment means the ‘rich get richer’ more so. However, there is a limit to how rich you can get – any given locus can influence a maximum of all \( N \) other loci. As \( \alpha \rightarrow \infty \), preferential attachment causes all epistatic influence to be ‘bunched up’, to be allocated to a minimal number of maximally-influential loci, leaving all other loci merely influenced. All the power is in the fewest number of hands. In practice we find that maximally-unequal distributions are produced at \( \alpha \approx 3.0 \), and further increasing \( \alpha \) merely reduces the noise inherent in the stochastic preferential attachment process. Therefore, the interesting range of \( \alpha \) is the one we have explored: between \( \alpha = 0 \) (no preferential attachment reproduces Kauffman’s results for the random scheme) and \( \alpha \approx 2.5 \), on the verge of a class of indistinguishable, maximally-unequal topologies. In this ‘interesting in-between’, preferential attachment creates novel structure in the epistatic network, not found in the local or random schemes, and exploitable by the evolutionary process in exploring the resulting landscape.

For both the local and random variants of the \( NK \) model the correlation between the fitness of a local optimum and the fitness of its neighbours decreases with increasing \( K \). Figure 4.8 compares the distribution of fitness values of genotypes adjacent to both a randomly selected point on the landscape and at a local optimum across a range of values of \( K, \alpha, r \). The comparison is made for \( N = 96 \) and \( K = 8 \), but the qualitative results are characteristic of the comparison in general. For the local and random variants, fitness values adjacent to a both random starting points and local optima are relatively tightly distributed around the selected point. This distribution is far more skewed for the \( NK\alpha \) variants, with many neighbours very highly correlated to the selected point, but a few highly uncorrelated neighbours, both at random points and at local optima. This distribution is strongly influences by the choice of \( \alpha \), with \( \alpha = 0 \) reverting to the same distributions found in the \( NK \) random case.

\(^3\)Provided by the GNU R wilcox.test() function
As an adaptive walk proceeds, we expect a number of measures to be related to the step number in some way. Figure 4.9 plots the relationships between the step number and mean fitness, the number of mutations that must be tested before a fitter neighbour is discovered, and the maximum degree of mutated loci. The first two of these measures are useful as a validation method – we know that fitness should increase as a walk progresses, so correlation between step number and mean fitness should be positive. We also know that as $K$ increases, walk lengths tend to be shorter and the optima discovered of lower fitness (even with our new scale-free topologies, if less so). These are uncontroversial, and borne out by the data. More interesting is the third subplot — ‘max degree’. For the local and random variants of the $N K$ model, loci are (roughly) epistatically equipotent. However, in the $N K \alpha$ model, some loci are more influential than others. How does this affect the rate at which different loci are mutated during an adaptive walk? At each step of an adaptive walk, recording the out-degree of the mutated gene (how many loci it influences epistatically), then plotting the correlation of this with the step number reveals that the most influential loci become fixed early in the adaptive walk. This relationship depends on a level of inequality in the locus out-degree distribution. When $K$ is high, inequality drops, and we see the relationship between max-degree mutations and step number fall.

Kauffman used the term ‘Massif Central’ to describe a global structure he discovered in landscapes with small $K$. It refers to the tendency for high-fitness local optima to be located in the vicinity of the globally known optimum, rather than being randomly distributed as is the case for high $K$. The inverse correlation between fitness and distance and its gradual erosion for higher values of $K$, shown in Figure 4.10 confirms this result for our implementation of local and random epistasis. When we consider the $N K \alpha$ variant, we find a similar but stronger relationship with many fit optima close to the global optimum. Unlike for the original model variants, for the $N K \alpha$ variant this relationship between optima fitness and Hamming distance is maintained for much higher values of $K$, most notably between $\alpha = 1.5$ and $\alpha = 2.0$.

Figure 4.12 demonstrates that, for low $K$, the random $N K$ model variant produces optima with a range of basin sizes and that there is a weak correlation between the fitness of an optimum and the size of its basin of attraction. Increasing $K$ destroys this correlation, as the landscapes become increasingly convoluted, and effectively isotropic. However, the $N K \alpha$ variants give rise to optima with a variety of basin sizes even for high $K$ landscapes, and here, optima fitness is strongly correlated with basin size. This accounts both for the fact that adaptive walks are taken on $N K \alpha$ landscapes tend to be longer than those carried out for equivalent landscapes from the local or random model variants, that they tend to terminate at optima of higher fitness, and that the same optima are visited by many walks — the basins of attraction are vastly increased under the $N K \alpha$ scheme.
4.3.1 Assortativity

The contribution of assortativity to the results, beyond those significant differences contributed by the $NK\alpha$ already described is minor, from the results we have collected. Whilst there is a clearly discernible difference in the network topology, the methods we use to measure the resulting fitness landscapes show that walk length and optimum heights are much the same as the corresponding raw $NK\alpha$ variant, with $r$ making little impact (see Figure 4.13).

Figure 4.11 shows that assortativity has only a minor effect on the Massif Central, over and above the significant change seen by introducing the $NK\alpha$ variant (Figure 4.10). Extreme values of $r$, where suitable dependency graphs are possible, slightly erode the correlation between distance and optimum fitness, but the relationship is still significantly elevated above the values found for the original local and random cases for the same level of $K$.

4.3.2 Epistatic structure and landscape structure

One of the most striking results we have found is the size and persistence of the Massif Central global feature of the fitness landscape, particularly under the $NK\alpha$ scheme. We now want to ask how this feature corresponds to the properties of the epistatic network — are important loci also important in optima on the fitness landscape? To answer this we take the set of all bitstrings representing the optima found on the landscape as the $m$ rows of a binary matrix $m \times N$ and calculate the Shannon information entropy $H_i$ for each column. Many of the highest optima are located close together and have alleles in common, so we might assume that, as in the $K = 0$ additive case, loci have a preferred allele, 1 or 0. Taken over all the known optima for a landscape, $H_i$ is a measure of the strength of this preference for one allele or the other, or whether no such preference exists, i.e., $H_i = 0.5$. Figure 4.14 demonstrates that, when $K$ is low, we have an almost bimodal distribution between very high and very low entropy — loci tend to be either fixed or totally random. As $K$ increases, and the Massif Central feature is eroded, the correlation between optima is eroded too, and the binary entropy per locus tends towards 1.0, i.e., random, mirroring the lack of global structure in the landscape. Again, this relationship between global structure is maintained at higher levels of $K$ in the $NK\alpha$ variants.

We find, unsurprisingly, that when there is global structure it is because of alleles shared between many optima, minimising the hamming distance between them. We might ask, then, whether the two networks (dependency and optima) are related via these results? When there is still some global structure in the fitness landscape, is there a correlation between low entropy and high out-degree, implying that important, high out-degree loci are more entrenched somehow? We calculate the correlation between the predictive utility ($|0.5 - H_i|$) and the out-degree over all loci for several variants (see Figure 4.15) and observe a consistent trend across all schemes — local,
random and $NK\alpha$: there is a weakly negative correlation for values up to $K = 8$, and a positive correlation above that threshold. However, this is not the strong signal we were hoping to see.

Thinking again about the Massif Central, we can imagine that the values we record for this structure could well be the result of several subsets of local optima, each with a cluster of common alleles maintaining a low hamming distance, but distinct from one another, and, taken over all the optima, these alleles from different clusters will tend to cancel one another out — masking the low within-group entropy at important loci. To test for this, we also perform a factor analysis on the optima matrix to discover whether there are higher order clusters of alleles co-occurring, and calculate a second Pearson correlation between the out-degree and the uniqueness values resulting from the factor analysis. If the structure exists, but is masked in this way, this should reveal a much stronger correlation between the variance unexplained by the common factors and out-degree, i.e., loci with the least entrenched alleles (highest uniqueness) have the lowest out-degree. This relationship is indeed shown in Figure 4.15, it is weak for the random scheme, and stronger as $K$ and $\alpha$ increase, providing an increasingly skewed out-degree distribution to work with.

So, there is little entropy associated with individual loci or alleles: the structure is higher-dimensional than that. It consists of groups of co-occurring alleles that are locked-in, entrenched — the ‘features’ of the landscape to navigate by. The structure is also hierarchical in a way: once an adaptive walk finds enough of the allelic combination for one of these clusters (they determine the large basins) the rest is mere details. An evolving solution would have to get the details really wrong to suffer too much with the strong fitness imbued by the core allele set.

The evolvability of changing a low-$K$, low-$N$ dependency network into a higher $N / K$ network and exploring a bigger space is difficult with uniformly distributed complexity, but if information about what already works reliably can be brought forwards (capturing information about the world, or at least the problem) then this increase in complexity can be managed and encapsulated: a kind of stratified stability (see Bronowski (1970)).
Figure 4.6: The distribution of walk lengths to an optimum is shown for adaptive walks on fitness landscapes generated for increasing $K$, and under epistatic topologies (uniform / local, random, scale-free with exponent marked ‘a’).

Chapter 4 Scale-free epistasis in the NK model
Figure 4.7:
The distribution of local optimum fitnesses discovered by adaptive walks on fitness landscapes generated for increasing $K$, and under epistatic topologies: uniform; local; random; scale-free with exponent $\alpha$ marked ‘a’).
Figure 4.8: An assay of neighbour fitness at selected points on different fitness landscapes.

Correlation between neighbours’ fitness determines landscape ruggedness. To illustrate this for a given point we plot its fitness as a horizontal line and the fitness of all $N$ 1-bit mutational neighbours as vertical lines (we ‘unroll’ a 360° panorama onto the page). For each landscape type we compare ruggedness at a local optimum (in red) and a random point (in green).

Diverse neighbour fitnesses, as seen for both points on the $K_{max}$ landscape, indicate low correlation, i.e., high ruggedness. By comparison, the majority of neighbour fitnesses for scale-free plots are tightly distributed around the central point indicating a smoother landscape. Notably, a few neighbours of the scale-free optima have significantly lower fitness, showing the presence of highly deleterious mutations.

We use the fitness of the neighbours at the random point as an index to order both sets of vertical lines (seen as a smooth increase left-to-right for neighbours of the random point in green). This ordering reveals that, in the scale-free cases, the severely deleterious mutants at local optima were also significant at the random point — indicating global structure in the landscape, and the possibility that sources of instability at local optima can be selected against much earlier in an adaptive walk, i.e., become locked in evolutionarily.
Figure 4.9: These plots show the correlation (Pearson correlation coefficient $r$) between walk step number, and three measurements that help validate our reasoning about the progress of adaptive walks in the varying landscapes.

First, we expect **mean fitness** to be positively correlated with walk step: even when walks are very short, we are using a hillclimbing algorithm – it can only go up!

Second, from Kauffman (1993), we expect the number of **mean mutants tested** at each step to increase: the more progress the hill-climber makes, the harder it becomes to make further progress.

Finally, mutations at high-degree loci (**max degree**) have the highest effects on fitness. We therefore expect such risky mutation events to be somewhat negatively correlated to walk step number – tending to occur early in a walk when they have much to add, and less so later in a walk when there is more to lose. This relationship ceases to hold when there is little inequality between loci (when $K$ is high), or when there is no inequality between loci at all (uniform ‘local’ epistasis – the blue line).

$p < 0.01$ for all points.
Figure 4.10: The ‘Massif Central’, a cluster of high peaks, is revealed by plotting the correlation between the fitness of local optima and the Hamming distance to the highest known point.

In the top-left plot, local with $K = 2$, the elliptical grouping shows a clear inverse correlation: as Hamming distance (from the fittest discovered optimum) increases, optima fitness falls. I.e., the highest peaks are clustered close together. Kauffman (1993) named this remarkable global structure the ‘Massif Central’.

For the local and random cases (top two rows) our results confirm those in the literature: the Massif Central is present when $K \leq 4$, and this relationship weakens as $K$ increases. Visually the grouping changes from elliptical to circular as $K$ increases. By $K = 8$, the relationship is nil – the distribution of optima is centred on mean fitness and mean Hamming distance ($\frac{N}{2}$) as we expect when sampling a random (anisotropic) landscape. Using our scale-free epistatic topology (lower three rows), we observe the persistence of global structure at higher values of $K$, even when $K = 8$, a significant new result. Data is aggregated from multiple such landscapes. The colour indicates how many optima exist in each ‘bucket’. This is not the number of times these optima are found during evolutionary walks (the basin size). For that measure, see Figure 4.12.
Figure 4.11: The Massif Central is not significantly affected by assortativity, and remains present at high levels of epistasis ($K = 8$), as for the untuned scale-free graphs shown in Figure 4.10.

Global structure is revealed by plotting the correlation between the fitness of local optima and the Hamming distance to the highest known point. Here we use our scale-free epistatic topology with $\alpha = 1.5$, and vary assortativity from dissortative to assortative ($r \in -0.3, -0.1, 0.1, 0.3$).
Figure 4.12: The sizes of basins of attraction, measured by the number of times an optimum is reached from random starting points, are plotted for different levels of epistasis $K \in \{2, 4, 8\}$ (arranged here by column), and the different epistatic topologies (arranged by row: Kauffman’s local, random, then our scale-free with $\alpha \in \{1.5, 2.0, 2.5\}$). The number of hits is a logarithmic scale, and the density is given using a colour from the red to blue spectra heat scale.

Kauffman noted a positive correlation between hits and fitness when $K = 2$, seen as a trend from bottom left to top right in the spread of points (in particular under random epistasis) but noted that this fell away with higher $K$.

We see that scale-free epistasis has a striking effect, increasing the basins of attraction such that many high fitness peaks are reached two or three orders of magnitude more often than before, even at high values of $K$, and that the correlation between basin size and fitness is maintained.
Figure 4.13: Comparing the length of adaptive walks and fitness of optima on landscapes with assortative and dissortative graphs. The mean length of adaptive walks, and fitness of optima for local, random, and scale-free graphs were aggregated from 100,000 adaptive walks for each combination of $(K, \alpha, r)$. In all cases $N = 96$. Note that for higher values of $\alpha$, there are often no results for extreme values of $r$, as such networks cannot be created. Broadly, we observe that assortative epistatic graphs (blue) tend to result in shorter walks to lower fitness optima.
Figure 4.14: Distribution of conserved individual alleles. We calculate the binary entropy for each locus across all optima found (red), and also across a subset of optima located within Hamming distance $N/3$ of the globally known optimum (blue). A low binary entropy value corresponds to a highly conserved allele across all optima. At low $K$ we observe bi-modal distributions – some loci with highly conserved alleles, other loci are effectively random. As $K$ increases, the highly conserved (low binary entropy) loci are lost, corresponding to the erosion of global features in the landscape. Our scale-free topologies do support more conserved alleles at slightly higher values of $K$ than Kauffman’s original local and random topologies (compare $K = 4$ top row to lower three rows). However none of these plots show significant numbers of conserved alleles at $K = 8$ which we might expect, in order to explain the global structure in these landscapes revealed by other measurements. Not shown are the $K = 0$ case, where all the loci at the single global optimum have entropy of 0.00, and the $K_{max}$ case where all loci are random and have entropy $\approx 1.00$. 
Chapter 4 Scale-free epistasis in the NK model

Figure 4.15:
Distributions of per-locus correlation between out-degree and predictive utility and out-degree and uniqueness.
The local model is not shown as it has uniform out-degree distribution, and therefore no variance.
The top row shows the random model and the scale-free model with $\alpha = 1.0$ as a baseline. The second row shows the raw scale-free model with $\alpha \in \{1.5, 2.0, 2.5\}$. The third row shows aggregated dissortative ($r < 0$) results for the same values of $\alpha$, for column-wise comparison, and the fourth final row shows the aggregated assortative results.
4.4 Conclusions

Broadly, fitness landscapes vary in size according to $N$, and correlation according to $K$. In detail, the structure of the network of $K$ epistatic connections has a significant effect on the global features of the resulting fitness landscape. Any global features of a fitness landscape represent structure which can be exploited by an optimisation process, such as evolution.

In general, imposing an increasingly scale-free structure on the network of epistatic interactions brings about a number of significant changes to the behaviour of adaptive walks on the associated fitness landscape: longer adaptive walks, larger basins of attraction, marginally higher fitness optima at higher $K$, more clustering of optima in the landscape and increased correlation between their fitness and the distance between them.

In previous published results for the $NK$ model, as $K$ goes to $N - 1$, the landscape becomes more convoluted and statistical properties in one part of the landscape are largely predictive of the whole. Consequently, the effect of increasing $K$ is to impose ruggedness globally — the complexity catastrophe. A small amount of epistasis ($K = 2$ or $K = 3$) has beneficial effects, allowing the system to choose from a greater number of possible allelic combinations (because of the linkage) without being overwhelmed by conflicting constraints. When the same number of epistatic interactions are allocated non-uniformly, the genome is structured such that there exist a few influential loci and a majority of loci with little influence. This structure gives rise to similar global structures that we have seen in the random and local cases at low $K$, but allows these features to persist even at middling values of $K$ — the non-uniformity allows some mitigation of the complexity catastrophe, and massively enhances some properties of the landscape, for instance the sizes of basins, ensuring that many adaptive walks reach the same solutions. Portions of the landscape exhibit properties that are very different from one another. More specifically, fixing alleles at influential loci confines an adaptive walk to a relatively correlated sub-landscape, while fixing the same number of low-influence loci confines an adaptive walk to a much less correlated landscape. Adaptive walks on such landscapes tend to initially spend time fixing influential loci, since mutating these alleles can bring about significant fitness changes. Once a satisfactory configuration of highly influential loci is discovered, low influence loci can be fixed relatively easily, since each is essentially independent from the others.

The Roman philosopher Seneca noted that if you have more to lose than to gain from fate, then this isn’t a happy asymmetry. Seneca’s comment seems relevant to our model — we observe in adaptive walks on correlated landscapes that the higher the out degree of a locus, the less frequently is it mutated as fitness increases. In other words, the most connected loci has the greatest knock-on effects by changing the fitness contribution of other loci as well as themselves, and the greater the number of changes, the greater the overall change in fitness there is. As this change is random, there are two observations to be made.
First, the landscape is only correlated in as far as the mutation changes between two neighbours in space are small changes — i.e., the fitness at one point is correlated to the fitness at its neighbours. Occasional large changes are permitted, but if the majority of fitness differences between neighbouring points in geno/pheno-space are large fitness differences, we find ourselves not in a correlated landscape but a rugged one, in which there is little gradient hint for an adaptive process to follow.

Second, there is an asymmetry as the fitness increases — if there is a relatively large fitness gain to be had by mutating the allele at a highly dominant locus, once it has been exploited, as we get fitter and fitter there is more fitness to lose by mutation than there is to gain. As we near the top of a peak, the number of ways ‘up’ are vastly outnumbered by the number of ways down. Conversely, at the bottom of the valley almost all directions lead to an improvement in circumstances, and those neighbours that involve mutating a high-degree locus are the least correlated to the current fitness, echoing Kauffman’s ‘long jump’ adaptations whereby early in an adaptive walk it makes sense to make large, random jumps in the landscape — a gamble, but one that is more likely to pay off from a mediocre starting point. Rather than large, multi-loci mutations like this, the \( NK\alpha \) scheme makes such uncorrelated jumps available as immediate neighbours, which changes the shape of the landscape.

As all the loci are mutated to mutually coherent alleles, they become self-reinforcing — each new mutant at position X improves the overall fitness for those alleles already present. On the subsequent steps, if another mutant allele at locus Y improves fitness sufficiently to be adaptive, whilst also diminishing the contribution of the allele at X, then X may be mutated again later on.

While it has been known for some time that, for instance, the network of protein-protein interactions for yeast exhibits a scale-free degree distribution (Fernández, 2007), recent work has shown that although the network for ancestral yeast has high degree proteins tending to interact directly with one another, the network for contemporary yeast is less assortative, with what has been interpreted as a more modular structure (Gavin et al., 2006). For instance, precursors to modern yeast feature an epistatic network with a single hub related to the ribosome, whereas the modern yeast network exhibits two hubs, one ribosomal and the other related to signalling. These hubs are connected, but only via other poorly connected proteins, making the whole network appear modular (Fernández, 2007). Scale-free network topologies tend to be robust to failure unless the hubs are targeted (Albert et al., 2000; Barabási, 2003; Jeong et al., 2001), and a modular topology has the advantage of preventing the failure of one hub triggering the failure of another.

Another analysis of protein expression in yeast (Pál et al., 2001) finds that highly expressed genes evolve slowly in yeast. An explanation for this is that the prolific protein products of these genes are widely used throughout the organism, and so high pleiotropy reduces the number of possible loci where neutral mutations are available.
Assessing the effects of systematically tuning assortativity in our results, the strong qualitative change expected by forcing the dependency network to be dissortative is not as pronounced as the change introduced by the raw $NK\alpha$ scheme. The fact that assortativity is clearly a selected feature of natural networks, and yet does not have great impact here may be due to the way in which the adaptive walks proceed. Assortativity, specifically slightly negative assortativity by which hubs are only peripherally connected by low-degree nodes creating a modular topology, may come into its own when crossover is at play, i.e., modules of highly intra-dependent loci (with weaker interdependence between groups) are well adapted to some feature in the world and are selected as a group, or it may be that the $NK$ model landscape does not exhibit the kind of hierarchical problem structure that a modular topology appears to be so well suited to in the real world.

The human genome project revealed a far lower number of genes than anticipated, increasing the significance of the study of their interactions. Similarly, by extending an existing model, the current chapter demonstrates how a scale-free epistatic network topology alters the properties of a fitness landscape in a way that makes adaptive dynamics on it much more liable to discover high-fitness optima despite strong epistasis.

The key point is that the landscape has been altered by changing a property of the dependency graph topology. As $K$ increases, the basins shrink and the average optimum decreases in fitness. However, we can mitigate this effect by increasing the non-uniformity of the out-degree (increasing $\alpha$), and again, further, by lowering the assortativity which has the effect of making the dependency graph modular — there is a gradient of importance to solve both globally and per module, but this is manageable because of the fractal-like gradients. Relatedly, this fractal-like gradient in the importance of loci enhances an entrenchment effect — the most important parts of the network, those that effect the most other fitness components, tend to fix earlier during adaptive walks. In fact, we can see sub-module optima (groups of alleles that frequently co-occur in optima) that go to make up whole-genome optima in the factor analysis.

This ability to break down a problem into sub-problems of different sizes, with an implicit gradient of contribution and therefore a given ordering, allows problems with greater interdependencies to remain tractable via evolutionary methods. The structuring of the dependency graph in such a way as to remain evolvable is itself an evolved trait.

It is also here, finally, using the uniqueness vector from the factor analysis and the out-degree, that we observe a signal that has the properties that suggest it will be preserved over evolutionary time. An allele, a particular variant of a trait (at a locus) may by itself not be a reliable signal for a process to usefully latch onto. When we considered all the optima, and subsets thereof, looking for biases by locus, none was found. However, when we reduce the dimensionality of the space via factor analysis, we reveal that there are clusters of alleles that co-occur and that these clusters are significantly correlated to the dependency network structure. So in this abstract
fitness landscape, we cannot say that allele \( a \) is better than \( b \), that it appears more frequently in local optima. We can say that, having explored the landscape, there exist high correlations between alleles, particularly at high-degree loci, and that these represent entrenched features.

Entrenched features can be exploited as they are hard or impossible to mutate to alternatives. It is exactly this sort of entrenchment that is our bridge to models of immune system coevolution: an entrenched feature in hosts is something that pathogens can exploit, and conversely an entrenched feature in pathogens is the sort of thing hosts may be able to use in order to identify them, and that they will find hard to evolve away from, even given a huge speed advantage in terms of generational times.
In this thesis, we opened by asking why questions about function and the evolution of immune systems. Why do immune systems exist at all? Why are they so ubiquitous? We reviewed the literature on immunology, and on the related field of Artificial Immune Systems research. We defined our particular interest in functional questions about the evolution of immune systems, in looking beyond the how and why of the specific immune systems we observe in life-as-it-is, to broader questions of the generality of immune responses we might expect to see in life-as-it-could-be. Or, going further, in extracting the computational processes implicit in these systems to new problem domains altogether. We defined three research questions on the relationship between epistatic constraints, evolvability and lock-in, deemed critical for immune system recognition.

5.1 Summary of the work and results

In reviewing the literature on Artificial Immune Systems in Section 2.5, we considered how our work would fit within the three schools identified by Cohen (2007), and concluded that there is in fact a fourth school, centred around meta-modelling of the immune system, i.e., understanding the best practices for extracting appropriate computational processes and representations from a natural system. We recognise a small but growing quantity of literature in this area. The identification of this fourth school is itself a contribution of the thesis.
We introduced Kauffman’s well-known NK fitness landscape model in Chapter 3 and further specified the hypotheses that structured epistasis would promote evolutionary lock-in and enhance evolvability compared to benchmarks in the literature.

In Chapter 4 we used our implementation of the NK model to run a large number of adaptive walks exploring a variety of fitness landscapes. We validated our model using Kauffman’s original local and random epistatic topologies, and contrasted and compared these published results to those from our new scale-free and assortative topologies. The significant finding and contribution from this work was that the ‘complexity catastrophe’ produced by increasing epistasis can be mitigated by structuring the epistatic network in a way which is not arbitrary but is motivated by natural epistatic networks. Specifically, the Massif Central global structure and the basins that drain into it continue to exist at much higher values of K than previously possible. This supports the hypothesis (under RQ3) that a scale-free epistatic network would create fitness landscapes that are more accessible to evolutionary search at comparative levels of epistasis, and opens up problems at higher levels of epistasis that were inaccessible previously using uniform or random epistatic topologies. We acknowledge that, at lower levels of epistasis, a quantitative comparison of optima fitness scores actually places scale-free landscapes very slightly below those for less structured epistatic networks. However, a qualitative view of the landscapes favours scale-free epistasis on all other measures.

Connecting biologically-inspired epistasis and lock-in (addressing RQ2) was a tale of two parts. We found the incontrovertible evidence of lock-in during adaptive walks we expected. High-degree loci that have significant fitness influence tend to be mutated early (when there is more to be gained from an uncorrelated fitness change) and then fix, as making uncorrelated fitness changes later in an adaptive walk is far more likely to be deleterious (there’s more to lose). The second hypothesis under this question proposed that these locked-in alleles would be common to many of the high optima on the landscape – that assessing the entropy per-locus would reveal global preferences for particular locked-in alleles. This was not the case, but we still observed the Massif Central global structure, so there were certainly many common alleles shared between high optima, even if they weren’t all the same ones. To reveal the source of this structure, we had to perform a dimensional reduction (via factor analysis) to reveal that the structure was composed of ensembles of co-adapted alleles. These locked-in groups were not the single alleles we expected, but support the hypothesis that scale-free epistasis creates highly conserved features in the landscape, just at a higher-order than anticipated.

In reverse order we reach RQ1 “Is there a class of problem space where evolutionary lock-in can provide slower evolving hosts sufficient information to identify faster evolving pathogens?”. We can answer yes — the fitness landscape (problem spaces) resulting from the use of scale-free structured epistasis exhibited both locked-in features during adaptive walks, and at a global level (albeit groups rather than individual alleles as we had expected). These landscapes also had
evolutionarily beneficial properties: large basins of attraction surrounding high optima, even as the level of epistasis increased.

5.2 Further work

In related publications, Hebbron et al. (2009) and Hebbron and Noble (2009), we have made (somewhat abortive) attempts to situate the kinds of underlying dynamics we discovered here with the NK model into a substrate that could support a simulated agent-based ecology of hosts and pathogens. It is ambitious to try to contain the varying levels of abstraction in the same model (Levins (1966) would take issue). However, there is certainly value in attempting to put the theory into practice and see whether a coevolutionary engagement would fall into locked-in traits, and exploit them.

A more immediate and narrow piece of work would be to perform some meta-optimisation on the NK model as we have it. Currently, we have a good assay of the landscapes at specific points in \((N, K, \alpha, r)\) space, but where are the ‘most evolvable’ landscapes found? From the point of view of an adapting agent, forced to bear an epistatic load of \(N \times K\) edges, what values of \(\alpha\) and \(r\) should you chose to structure your epistatic network for the best fitness landscape?

Our investigations into the affect that assortativity in the epistatic network had on the landscape did not demonstrate any great departure from the scale-free results already found. Whilst many natural scale-free networks appear to be tuned to certain levels of assortativity, and we assume this is to some end, no evolutionary advantage was apparent from our experiments. We consider that as assortativity appears to be conserved across many natural networks, it does play a role, but in some way that the NK model does not obviously reveal, perhaps in maintaining well co-adapted modules of alleles during sexual reproduction. Further development of our NK model, incorporating alternative neighbourhood operators (e.g., crossover, rather than single-bit-mutation) might provide insight into the apparent selective advantage that this property has in real networks.

5.3 Theoretical insights

The significant corollary to these results is that certain allelic combinations become deeply entrenched – common to many of the high optima, strongly suggesting that such entrenchment is a phenomenon of all systems that evolve on correlated, but rugged landscapes, as biotic life must do.
The discovery of these higher-order locked-in features teaches us an important lesson. A naive immune algorithm that only exploited individually conserved alleles as the features of interest would fail to recognise pathogens based on these landscapes. The appropriate features of interest were combinations of alleles, which emphasises the criticality of problem representation. It is necessary for there to be some structure in a problem-space for there to be any possibility of efficient search, but not sufficient. The problem structure has to be reflected in the search algorithm itself, it must incorporate an appropriate representation of the problem space – a perspective that enhances features and ignores noise.

We note that the world (or umwelt) our own immune systems ‘sees’ has had to move from integrating information at evolutionary speed (the innate system) to be able to learn in almost-real-time to match the evolutionary speed of the pathogenic world it is exposed to. When we apply learning or optimisation algorithms, from machine learning, or from immune or other biologically-inspired sources, the most difficult part of the problem, the autonomous and adaptive part of learning, is in choosing suitable features by which to discriminate the space, to devise a map of the territory of possible inputs, and thereby navigate and select from a lower-dimensional set of outputs an appropriate (adaptive) response. In AIS, attention has been drawn to the use of shape-space as originally proposed by Perelson and whether it is really an appropriate feature detector, or merely a convenient computational abstraction (McEwan and Hart, 2011). This choice of features determines the properties of the fitness landscape, the perspective on the underlying problem that can determine whether it is algorithmically simple to solve, or so convoluted (with a bad choice of features) as to be intractable.

We have cited Wolpert and Macready (1997)’s “no free lunch” theorem a number of times, noting that it is not possible to create a good general purpose optimisation algorithm over all possible problem spaces. However, there is growing interest in the idea that whilst we may not be able to create a true general purpose optimisation algorithm for all possible problems, it may be possible to discover general algorithms for large swathes of problem-spaces that actually interest us (Koehler, 2007).

Mountcastle (1978) introduced his single algorithm theorem: neurons in the brain aren’t all that different from one another, even regions of the neocortex that we associate with special purposes, e.g. visual processing, can be turned to other tasks, as demonstrated in ferrets — visual inputs were rewired into the auditory cortex which quickly adapted to perform a similar function (orientation of the visual field) as would be expected to occur in the original visual region Sharma et al. (2000). In humans too, Bach-y Rita et al. (2003) demonstrated a widely reported device for ‘seeing’ the world through a matrix of electrodes on the tongue, fed from a video camera — subjects were quickly able to make use of the signal as visual input despite the unusual route, again demonstrating both the plasticity of the brain, and suggesting that it is the structure of the signal that determines how the neurons organise themselves, rather than the other way around,
that specialised neurons are specifically prepared in such a way as to process signals in a particular
domain.

If Mountcastle is on the right track, that there is a single algorithm running on all the neurons
in the neocortex, learning and adapting, modelling the world and adapting that model as new
information comes in, then this is the goal of much of AI and computer science. If there is such
an algorithm, a general purpose learning algorithm that discovers and maintains its own feature
extraction as well as multiple hierarchies of abstraction and compression above the raw data, it is
not inconceivable that, having chanced on this algorithm at least once, a variant underlies the
learning and adaptation in the immune system — a deep organising principle of information as it
occurs in the real world.

What might these deep organising principles be? They must include the processes of self-
organisation and of evolution through natural selection on somewhat rugged, somewhat correlated
landscapes that biology has been exploring and learning from since life got started.


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Linnæus, C. (1758). Systema naturae per regna tria naturae :secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Impensis Direct. Laurentii Salvii.,


REFERENCES


REFERENCES


